The dissertation committee for Damien Sean Eldridge certifies that this is the approved version of the following dissertation:

Essays in Microeconomic Theory

Committee:

Maxwell B. Stinchcombe, Supervisor

Stephen G. Donald

Kenneth Hendricks

David S. Sibley

Andrew B. Whinston

Thomas Wiseman

Essays in Microeconomic Theory

by

Damien Sean Eldridge

B.Ec., B.Sc., Grad.Dip.Ec., M.Ec., M.Sc.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2007

This dissertation is dedicated to my family: Mum (Robyn), Dad (John) and my siblings (Lana, Mark, James, Luke, Mikaela, Ashleigh, Marissa and Zac). It is also dedicated to the memory of my grandparents: Nan Franklin (Roma), Pop Franklin (Jim), Nan Eldridge (Norma), Pop Eldridge (Bill) and Great Nan (Agnes).

Acknowledgements

There are a number of people that I would like to thank for the time and effort they have devoted to helping me complete this dissertation, as well as for the patience they have displayed along the way. First and foremost, I would like to thank my supervisor, Max Stinchcombe. In addition to supervising my dissertation, Max actually taught me on three separate occasions. He clearly has the patience of a saint! The other members of my dissertation committee; Stephen Donald, Ken Hendricks, David Sibley, Andrew Whinston and Tom Wiseman; also deserve a great deal of thanks for taking the time to read my dissertation and providing helpful advice. In addition to my dissertation committee, I would like to thank Preston McAfee for his helpful advice during the early stages of my dissertation research. I was also fortunate to receive helpful comments at the Australasian Economic Theory Workshop in February 2006, as well as in seminars at the Australian National University, the University of Melbourne, La Trobe University and the University of South Australia. In this regard, I would particularly like to thank Peter Bardsley, Suren Basov and Xiangkang Yin.

Dealing with administrative trivia is rarely pleasant. When you live in a different hemisphere to the administrative offices, this process is even more frustrating than usual. In these circumstances, a sympathetic graduate coordinator is a blessing. Vivian Goldman-Leffler is such a person. In her able hands, the tyranny of distance was reduced to merely a mild form of despotism!

The fact that I live in Australia increases the cost of defending this dissertation. In addition to the time and effort that is involved in preparing for the oral defence, there are significant travel and accommodation costs. As such, I would like to acknowledge the financial support of my current employer, La Trobe University, in the form of a start-up research grant. The funds from this grant will defray most, but not quite all, of these travel and accommodation costs.

I would also like to thank a number of my friends for their encouragement and support. They made the journey much more enjoyable than would otherwise have been the case. These friends include John Barry, Rafi Berkson, Lisa Blanton, Natalya Brown, Peter Chang, Allen Fawcett, Anne Golla, Caitlin Knowles-Myers, May Li, Martha Martinez Licetti, Diya Mazumder, Ilke Onur, Carmen Ponce San Roman, Justine Power, Matthew Power, Christy Spivey, Malathi Velamuri, Travis Warziniack and Elaine Zimmerman.

More than anyone, I would like to thank my family for their continuing support. It is sometimes said that you can choose your friends but not your family. As such, I am very fortunate to have been blessed with such a wonderful family! Their emotional and financial support throughout my life has allowed me to become the person that I am. They deserve the credit for anything good that I do. I alone bear the blame for everything else!

There is one other person who deserves a great deal of thanks. This person is Br Kevin Herlihy FMS. Br Kevin devoted a great deal of his life to teaching many young people, including me, many important things. These include leadership, bush-craft, survival skills, public speaking skills and the importance of community service. The most important thing that Br Kevin encouraged in the students he taught was a belief in themselves and their own innate goodness.

In many ways, he has been a mentor to me. I wish I had learnt the lessons he taught much better than I actually did. Especially the one about being gentle on myself.

Finally, I would like to pay tribute to someone who I suspect wanted me to achieve this degree more than anyone. This is my late grandmother, Roma Franklin. I am sorry that I couldn't complete my dissertation in time for you to see it.

Essays in Microeconomic Theory

Publication No. ______________________

Damien Sean Eldridge, Ph.D.
The University of Texas at Austin

Supervisor: Maxwell B. Stinchcombe

My dissertation uses game theoretic techniques to explain the existence of two economic institutions which are ignored by the Arrow-Debreu-McKenzie general equilibrium model of an economy. The first institution that I consider is the gated structure of some professional service industries. Two very similar explanations for the existence of such a structure are provided in Chapter 2 and Chapter 3 of this dissertation. The second institution that I consider is the presence of self-enforcing social conventions that can allow a small, isolated village to successfully manage common property resources in the absence of private property rights or some other form of explicit regulation. An explanation for the existence of this institution is provided in Chapter 4 of this dissertation.

Many service industries, including the medical and legal professions in some countries, display a gated structure. Rather than approaching a final producer directly, a consumer will first seek a referral from an intermediary. Chapter 2 provides one possible explanation for such an industry structure. If the outcome

of a transaction depends on producer effort, which is unobservable and unverifiable, then the market may fail to generate a Pareto optimal outcome. This is the standard moral hazard problem. If consumers had a long-run relationship with producers, this type of market failure might be avoided. However, in some industries, consumers will only have a short-run relationship with producers. A gatekeeping intermediary may provide an opportunity for reputation effects to apply in such a setting. By aggregating many potential consumers, gatekeeping intermediaries can create an artificial long-run relationship between a consumer and a producer. This long-run relationship reduces the incidence of shirking on the part of the producer.

Chapter 3 provides another possible explanation for the gated structure of some professional service industries. Such an industry structure might help to alleviate adverse selection problems between parties that interact infrequently. Intermediaries aggregate many short-run transactions between various consumers and a particular producer. As such, they might be able to learn a producer's level of proficiency more rapidly than an individual consumer. However, the presence of a positive information externality means that too few consumers will seek a referral. As such, some form of regulation to encourage consumers to seek a referral might be warranted.

Chapter 4 provides a model in which small and relatively isolated communities can successfully manage local commons informally in circumstances where larger or less isolated communities could not do so. The reason for this is the non-anonymous nature of many interactions between the members of a small and isolated community. Such communities may be able to use these multiple interactions to enforce informal restrictions on the usage of local commons. To the extent that the process of economic development reduces the number of non-anonymous interactions among community members, it will reduce the ability of the community to successfully manage the local commons informally. The resulting need for either explicit regulation or the introduction of private property rights represents a hidden cost of development.

Contents

List of Tables **xiv**

List of Figures **xv**

1 Introduction and overview **1**

2 A shirking theory of referrals **4**

2.1 Introduction 4

2.2 Motivating examples 6

2.3 A competitive model of health care markets 9

2.3.1 The timing of the market in each period 10

2.3.2 Player objectives 12

2.3.3 The nature of equilibria 17

2.4 The patient-specialist supergame 17

2.4.1 The treatment choices of specialists 19

2.4.2 The treatment choices of patients 24

2.4.3 Long-run and short-run relationships 27

2.5 The GP-specialist supergame 29

2.5.1 Optimal deviation by a specialist 30

2.5.2 Incentive compatibility constraints for specialists 33

2.5.3 Participation constraints for specialists 39

2.5.4 Participation constraints for GPs 40

2.6 Industry structure with exogenous prices 41

2.6.1 The referral choices of patients 41

2.6.2 The equilibrium industry structure 44

2.7 Conclusion 46

2.8 Appendix: Industry structure with endogenous prices 55

2.8.1 Price formation without GPs 56

2.8.2 Price formation with GPs 57

2.8.3 Market outcomes with endogenous prices 59

3 A learning theory of referrals 61

3.1 Introduction 61

3.2 A competitive model of health care markets 63

3.2.1 The timing of the market in each period 65

3.2.2 Player objectives 66

3.3 Static outcomes in competitive health care markets 70

3.4 Dynamic outcomes without general practitioners 73

3.5 Dynamic outcomes with general practitioners 81

3.6 A comparison of treatment market outcomes 91

3.7 Information externalities and the need for regulation 92

3.8 Conclusion 97

4 Multiple interactions and the management of local commons 99

4.1 Introduction 99

4.2 A simple model of a village economy 100

4.2.1 A two person example of the village economy game 103

4.2.2 A three person example of the village economy game . . . 104

4.3 Outcomes in a village economy 106
4.3.1 The outcome for a short-run association of hermits 106
4.3.2 Outcomes for a short-run village 107
4.3.3 Outcomes for a long-run association of hermits 108
4.3.4 Outcomes for a long-run village 109
4.4 The impact of economic development 112
4.5 Conclusion 115
4.6 Appendix: A formal model of a village economy 117

Bibliography **119**

Vita **134**

List of Tables

2.1 Market outcomes 45

4.1 Payoffs in the common property resource interaction 101

4.2 Payoffs in other interactions 102

4.3 Two player common property resource interaction 103

4.4 Two player other interactions 103

4.5 Three player common property resource interaction 104

4.6 Three player other interactions 105

List of Figures

2.1 Market outcomes when $\theta B < \frac{C}{\delta} < \widehat{r}_1 < B$. 47

2.2 Market outcomes when $\frac{C}{\delta} < \theta B < \widehat{r}_1 < B$. 48

2.3 Market outcomes when $\frac{C}{\delta} < \widehat{r}_1 < \theta B$. 49

3.1 The relationship between threshold treatment prices 92

Chapter 1

Introduction and overview

Ever since Adam Smith introduced the notion of the invisible hand[1], economists have explored the ability of markets to improve the welfare of every member of society. This line of research culminated in the Arrow-Debreu-McKenzie general equilibrium model of an economy.[2] The notion of the invisible hand, along with a set of circumstances in which it is guaranteed to apply, is formally captured in the general equilibrium model through the first fundamental theorem of welfare economics. The first fundamental welfare theorem states that, under certain circumstances, the equilibrium of an economy will be Pareto efficient. An outcome is Pareto efficient if, and only if, there is no alternative feasible outcome that would improve the welfare of at least one individual without harming anyone else. The basic idea that underlies this theorem is very simple. Voluntary exchange between two individuals cannot harm either party and will probably

[1] The concept of the invisible hand was introduced to economics by Adam Smith in his books *The theory of moral sentiments* (Smith [109]) and *An inquiry into the nature and causes of the wealth of nations (*Smith [110] and [111]). An interesting discussion of these books can be found in O'Rourke ([78]).

[2] Useful overviews of the Arrow-Debreu-McKenzie general equilibrium model of an economy can be found in Border ([18]), Mas-Colell et al ([68], Part 4, Chapters 15-20), Starr ([115]) and much of Takayama ([122]). The seminal references for this model include Arrow and Debreu ([9]), Debreu ([29]), McKenzie ([69], [70], [71]) Negishi ([76]) and Nikaido ([77]).

improve the welfare of both parties. Otherwise, the party that was harmed would not agree to the exchange. As such, if nobody who is not involved in the exchange is harmed, the outcome of voluntary exchange will be a Pareto improvement.

A market failure is any situation in which the completely free operation of markets does not result in a Pareto efficient outcome. A necessary condition for a market failure to occur is that at least one of the assumptions underlying the first fundamental theorem of welfare economics be violated. These assumptions include market completeness, price taking behaviour by all economic agents and locally non-satiated preferences for all economic agents. The main types of market failure are various forms of imperfect competition, various forms of externalities and various forms of asymmetric information.[3]

While the general competitive equilibrium model of an economy provides an important benchmark, it leaves many interesting questions unanswered. In particular, it does not provide an explanation for the existence of many economic and social institutions. The failure of the Arrow-Debreu-McKenzie model to explain the existence of institutions provides the motivation for the topics addressed in this dissertation. My dissertation uses game theoretic techniques to explain the existence of two economic institutions which are ignored by the Arrow-Debreu-McKenzie general equilibrium model of an economy. The first institution that I consider is the gated structure of some professional service industries. Two very similar explanations for the existence of such a structure are provided in Chapter 2 and Chapter 3 of this dissertation. The second institution that I consider is the presence of self-enforcing social conventions that can allow a small, isolated village to successfully manage common property resources in

[3] A detailed exposition of the welfare properties of Walrasian equilibria in the Arrow-Debreu-McKenzie general equilibrium model of an economy is provied in Chapter 16 of Mas-Colell, Whinston and Green ([68]). The seminal papers on the two fundamental welfare theorems in the context of the Arrow-Debreu-McKenzie model are Arrow ([5]) and Debreu ([27], [28]).

the absence of private property rights or some other form of explicit regulation. An explanation for the existence of this institution is provided in Chapter 4 of this dissertation.

The presence of various market failures provides a potential role for the existence of the economic institutions that are considered in this dissertation. The market failure that is considered in Chapter 2 is a moral hazard problem. The market failure that is considered in Chapter 3 is an adverse selection problem. The market failure that is considered in Chapter 4 is an externality problem. In all three cases, the market failure occurs because of the short-run nature of the relationships between the various parties to a transaction. If those parties had a long-run relationship, then they could overcome the market failure through the use of reputation effects and the threat of punishment in the future. The role of the institution is to provide an artificial long-run relationship between the various parties to a transaction.

Chapter 2

A shirking theory of referrals

2.1 Introduction

The potential for moral hazard problems to result in market failure is well understood.[1] Indeed, there is a large literature on the design of contracts to alleviate moral hazard problems.[2] This literature focuses on a static setting in which the principal and the agent interact only once. Except in very restrictive circumstances, the market failure can at best be only partially mitigated. For this reason, moral hazard and other problems involving asymmetric information are often used to justify a variety of consumer protection related regulations.[3] But such regulations are potentially costly and sometimes ineffective.

[1] See, for example, Arrow ([6], [8]) and Pauly ([81], [82]).

[2] Useful surveys of this literature are provided by Hirshleifer and Riley ([49]), Laffont and Martimort ([62]), Macho-Stadler and Perez-Castrillo ([64]) and Mas-Colell et al ([68]). The central references include Arrow ([6]), Grossman and Hart ([44]), Hermalin and Katz ([48]), Holmstrom ([52]), Jewitt ([55]), Mirlees ([72]), Rogerson ([94]), Ross ([95]), Shapiro and Stiglitz ([105]) and Shavell ([106], [107]).

[3] See, for example, the discussions in Damania and Round ([26]), Hadfield et al ([45]) and Smith ([112]).

In many settings involving moral hazard, the transacting parties interact more than once. Repeated interaction potentially allows for greater alleviation of moral hazard than is possible in a static setting, through the use of dynamic punishment strategies and reputation effects.[4] Unfortunately, there are also many occasions in which parties to a transaction do not repeatedly interact. In the absence of repeated interaction, reputation cannot be relied upon to deter moral hazard. We might expect moral hazard problems to be rampant in markets characterised by few or infrequent interactions between the trading parties. Indeed, Mooney and Ryan ([73], p. 134) raise exactly this concern in relation to health care markets:

> "Whilst it is possible that repeated interactions constrain doctors' behaviour, since doctors will not want patients to lose faith in them, such repeated interaction will not take place in all sectors of the health care market. Whilst a dynamic model may be applicable to the GP-patient interaction (since there will be repeated interactions), it is less clear how applicable such a model is to the specialist-patient interaction."

Is consumer protection regulation the only safeguard available in such settings or can institutions be devised that might capture the benefits of a long-run relationship? This paper explores one possible solution. It involves the creation of intermediaries that generate an artificial long-run relationship between the transacting parties by aggregating many short-run relationships. In effect, the intermediaries act as a surrogate long-term partner, leveraging their own repeated relationship with the two transacting parties. This allows the short-run

[4] See, for example, Abreu et al ([1], [2]), Atkeson and Lucas ([10]), Fudenberg et al ([39]), Radner ([87], [88]), Radner et al ([89]), Rogerson ([93]), Rubenstein and Yaari ([98]), Spear and Srivastava ([113]), Stigler ([116], p. 179), Thomas and Worrall ([123]) and Townsend ([126]).

agents to build up a reputation for quality and the short-run principals vicarious access to that reputation.

2.2 Motivating examples

Variations of this industry structure can be found in many professional service industries, including the medical and legal professions. In many countries, these professions are organised around a gatekeeper. Access to the ultimate producer frequently requires a referral from an intermediary. In this paper, we will explicitly model the organisation of health care markets. However, the model readily translates into the organisation of some other service industries, including the legal profession.

In the medical industries of some Commonwealth countries, it is unusual for patients to visit a specialist without first obtaining a referral from a general practitioner (GP).[5] The GP is essentially the family doctor. He typically sees a patient many times throughout the patient's life, treating a variety of minor illnesses and referring the patient to an appropriate specialist for more serious complaints. As such, the patient and the GP interact repeatedly over a long period of time. Furthermore, because the GP has a pool of patients, he will typically encounter particular diseases many times. As such, the GP has the opportunity to develop a long-run relationship with particular specialists. While some severe or chronic complaints might require repeated interaction between a patient and a specialist as well, many patient-specialist relationships are inherently short-run. In such circumstances, the GP can potentially leverage his long-run relationships with both patients and specialists to induce an artificial

[5] Commonwealth countries in which many patients obtain a referral from a general practitioner before seeking the services of a specialist include Australia ([83], p. 421; [23], part 2, p. 3), New Zealand ([97], section 2, p. 14) and the United Kingdom ([19]). This arrangement does not necessarily apply to all medical specialties within these countries. For example, a patient would probably seek a referral before visiting an opthamologist in Australia, but would be unlikely to do so before visiting an optometrist.

long-run relationship between a patient and a specialist.

While the medical industry in the United States of America is not formally organised in the same way as it is in some Commonwealth countries, some of the key institutions in the US health sector enforce similar arrangements. Health maintenance organizations (HMOs) and preferred provider organizations (PPOs) combine health insurance and the provision of medical care.[6] HMOs typically require patients to see one of their gatekeeping medical practitioners before being referred to an approved specialist. By controlling which specialists receive patients and monitoring which patients are treated by a particular specialist, the HMO can effectively play a role similar to that of the GP in many Commonwealth countries. PPOs are essentially just a less restrictive form of HMO. They have no requirement for the patient to visit a gatekeeper before seeing a specialist. They do, however, provide financial incentives for patients to visit specialists on their list of preferred providers. By controlling which specialists are on this list, they can effectively punish specialists who are suspected of shirking. Just like the GP in our earlier example, HMOs and PPOs have a long-run relationship with both specialists and patients. They can leverage the financial clout provided by their relatively large customer base to punish specialists suspected of shirking.[7]

While we focus on the medical industry example in this paper, there are other industries with a gated structure that might in part be explained by this theory of intermediation, including the legal industry. The structure of the legal industry in some Commonwealth countries appears to be similar to that of the medical industry in those countries. A client needing legal services first visits

[6] Many undergraduate textbooks on health economics contain a discussion of HMOs and PPOs. A particularly good source is Folland et al ([38]).

[7] Clearly, HMOs and PPOs can punish specialists for a number of other undesirable behaviours too. Constraining costs by punishing suspected over-servicing may be one such concern. We are not suggesting that HMOs and PPOs exist solely to punish shirking by medical specialists. We are, however, suggesting that this is one of many roles that they can play.

a solicitor. If the service required is relatively minor, the solicitor may be able to take care of it himself. But if the client's case is going to trial, the solicitor might choose to brief a barrister, who will then represent the client at court. A client might have a repeated relationship with a solicitor because solicitors can handle estate planning, conveyancing and many other legal matters that do not require representation at trial. Even if a particular client does not have a repeated relationship with a solicitor, the client might have sought out the solicitor's services on the recommendation of a friend or family member who has previously employed him. A sequence of such recommendations can give rise to a sequence of clients that might be thought of as a single long-run client. A solicitor typically has many clients, so that repeated relationships between a solicitor and a barrister might also occur. In a similar fashion to a GP in the medical example, the solicitor may be able to leverage his long-run relationship with a particular barrister to ameliorate any moral hazard problem that might normally arise because of the short-run relationship between his client and the barrister.

As with the respective medical industries, the legal industry in the United States of America is not formally organised like it is in some Commonwealth countries. Once again, however, organisations have evolved that, among other things, play a similar role to solicitors in those Commonwealth countries. While many lawyers and some legal practices might specialise in a particular area of the law, multi-purpose law firms also exist. These firms are able to help their clients in a number of disparate and unrelated legal matters. With the exception of possible savings due to economies of scope arising from sharing fixed overhead costs, it is unclear why clients would not prefer to seek out different legal practices for different legal problems. The generation of an artificial long run relationship provides one further argument in favour of the use of multiproduct firms. By increasing the extent of interaction between particular clients and the legal practice, multi-purpose law firms provide the client with greater opportunity to punish poor performance. The amount of business the firm loses if the client drops them is greater than it would be if they were a single purpose practice.

2.3 A competitive model of health care markets

Consider an economy with three groups of agents who live forever. These groups are patients, general practitioners (GPs) and medical specialists. Let patients be indexed by $i \in \{1, 2, \cdots, I\}$, GPs by $j \in \{1, 2, \cdots, J\}$ and specialists by $k \in \{1, 2, \cdots, K\}$. We will assume that there are an infinite number of patients ($I \to \infty$) and specialists ($K \to \infty$), but only a finite number of GPs ($J < \infty$).[8]

In each period, a patient is randomly allocated a disease state ($d \in \{0, 1\}$). Patients may be either well ($d = 0$) or sick ($d = 1$). Following this, each sick patient can choose whether or not to seek treatment. Treatment can sometimes result in a cure, improving the patient's health status for that period. The probability that a disease is cured by treatment increases with the amount of effort the specialist devotes to the treatment. Patients can seek a referral to the specialist from a GP if they believe this will increase the probability that high effort treatment is provided. Both referrals and treatments come at a price. For budget constrained patients, the benefits of an increased probability of good health need to be weighed against the foregone consumption of other goods that expenditure on health care entails. We will assume that patients visit neither a

[8] The reason we assume that there are a countably infinite number of patients, a countably infinite number of specialists and only a finite number of GPs is that it allows us to use the standard version of the strong law of large numbers to make inferences about the number of sick patients in a GPs patient pool in each period. An alternative to this set of assumptions would be to assume that the number of patients is uncountably infinite, the number of specialists is countably infinite and the number of GPs is countably infinite. In order to analyse this alternative version of the model presented in this paper, we would need to use the techniques outlined in Judd ([57]).

GP nor a specialist when they are healthy.[9]

All agents in this economy are price takers who behave as though the existing prices are exogenously specified. We will focus on stationary equilibria for this economy, so that prices don't change over time. The price per referral from any GP is w, while the price per treatment from any specialist is r.

For payoff purposes, time is assumed to be discrete in this economy. Time periods are indexed by $t \in \{0, 1, 2, \cdots\}$, with payoffs occurring at the end of each period. In each period, the market opens and the agents interact within the market. Note that not all agents move at once in the market. The market process involves sequential moves by various agents. Thus the timing of the moves in the market process is important. We will maintain the assumption that time is discrete and index time within a period by $s \in \{0, 1, 2, \cdots\}$. In this fashion, each point in time can be given a unique time stamp of the form $(t, s) \in \{0, 1, 2, \cdots\}^2 = \mathbb{Z}_+^2$.

2.3.1 The timing of the market in each period

At the beginning of each period ($s = 0$), Nature randomly chooses a disease state for each patient, $d_i \in \{0, 1\}$. Each patient's disease state is chosen as an independent draw from some common distribution, $\Pi : \{0, 1\} \to [0, 1]$. The probability that any given patient is sick in any given period is π, while the probability that any given patient is well in any given period is $(1 - \pi)$. While each patient's disease state is private information, the distribution from which disease states are drawn is common knowledge.

At $s = 1$, having observed their disease state for the current period, patients choose whether or not to seek treatment. If patients choose to seek treatment, then they also choose whether or not to seek a referral from a GP. If they seek

[9] If the equilibrium prices for referrals and treatment are positive, this assumption is not needed. Even if these prices are zero, we could avoid making this assumption by introducing an opportunity cost of time (perhaps in the form of foregone leisure) into the model.

a referral, they choose which GP to visit at $s = 2$. If not, they choose which of the specialists that treats their disease type to visit. Recall that patients who are healthy are assumed to seek neither treatment nor referral.

We will assume that GPs follow up on the outcomes from treatment of any of the patients they refer. In this fashion, the GP knows the entire history of outcomes for each of his previous referrals at the start of each period. At $s = 3$, GPs choose the specialists to which they will refer their patients. Patients who seek a referral are assumed to follow the GPs advice and seek treatment from the specialist to which they are referred. We will assume that each GP refers all of his sick patients in a given period to the same specialist. This assumption is not essential. However, it does simplify the analysis when GPs have finite patient pools. The reason for this is that it allows specialists to estimate the size of a GP's patient pool from the number of patients that GP refers to him. The assumption is relatively innocuous when GPs have infinite patient pools, which is the case that we will focus on in this paper.

Following this, at $s = 4$, specialists choose how much effort to devote to treating each patient. For each patient, they can independently choose either high effort ($e = 1$) or low effort ($e = 0$). The effort choice affects the probability of the patient being cured.

Finally, at $s = 5$, Nature chooses whether or not each patient is cured. If a patient is cured, he will have good health in that period ($h = 1$), while if the patient is not cured, he will have bad health ($h = 0$). We will assume that high effort on the part of the treating specialist always results in the patient being cured, while low effort results in a cure with probability $\theta \in (0, 1)$. If the patient chose not to seek treatment, he will not be cured.

2.3.2 Player objectives

Every agent in this game is assumed to maximise the discounted present value of a sequence of per-period von Neuman Morgernstern expected utility functions. Furthermore, they all have a common rate of time preference, represented by the stationary discount factor $\delta \in [0, 1)$. Thus differences in the preferences of the three groups of agents arise from differences in their per-period preferences. These are outlined below.

Patients

Patients all have identical per-period preferences defined over their expenditure on health care (p) and their health state (h). These preferences may be represented by a quasi-linear per-period Bernoulli utility function of the form

$$u(h_t, p_t) = B(h_t) - p_t,$$

where $B(0)$ is normalised to zero and $B(1) = B > 0$.

The health state in each period is a random variable and may vary across patients. It depends on whether or not the patient is sick, whether or not treatment is sought and, if so, the effort devoted to treating the patient. Since each patient knows their disease status before having to make any decisions about treatment, the probability of good health in period t is given by

$$\gamma_t = \begin{cases} 1 & \text{if either } d = 0 \text{ or high effort treatment is received when } d \neq 0; \\ \theta & \text{if } d \neq 0 \text{ and low effort treatment is received;} \\ 0 & \text{if } d \neq 0 \text{ and no treatment is received.} \end{cases}$$

Note that the probability of good health depends on the amount of effort exerted by a specialist when treating a patient. This is not observed by patients, so that they do not know the exact probability of good health in period t after making their treatment and referral decisions. Let $\lambda_{i,t}$ denote a patient's belief that he will receive high effort treatment if he seeks a referral. In this paper, we will restrict our attention to symmetric equilibria in which each patient holds the same beliefs. Thus we can set $\lambda_{i,t} = \lambda_t$ for all $i \in \{1, 2, \cdots, I\}$. With these beliefs, each patient's subjective estimate of the probability of good health in period t is given by

$$\mu_t = \begin{cases} 1 & \text{if } d = 0; \\ \lambda_t + \theta(1 - \lambda_t) & \text{if } d \neq 0 \text{ and treatment is sought;} \\ 0 & \text{if } d \neq 0 \text{ and no treatment is sought.} \end{cases}$$

Expenditure on health care in any given period may also vary across patients. It will depend on whether or not the patient seeks treatment and, if so, whether or not the patient also seeks a referral. We will assume that patients do not seek a referral if they do not also desire treatment. Thus a patient's expenditure on health care is given by

$$p_t = \begin{cases} 0 & \text{if neither treatment nor referral is sought;} \\ r & \text{if treatment is sought without a referral;} \\ w + r & \text{if both treatment and referral are sought.} \end{cases}$$

Thus a patient's per-period expected utility is

$$Eu(h_t, p_t) = \mu_t B - p_t.$$

Patients do not know their future disease states. However, they do not have to make any decisions in any given period prior to observing their disease state in that period. As such, a patient's remaining lifetime expected utility after he has observed his disease state in any given period can be conveniently represented as:

$$U(p_t; d_t, H_{i,t}) = \mu_t B - p_t + \delta M,$$

where $H_{i,t}$ is the entire history that is observed by the patient prior to period t and M is the expected continuation payoff to the patient. The expected continuation payoff is the next period value of the total utility the patient expects to receive from all subsequent periods following the completion of the current period's stage game. Clearly the expected continuation payoff will be a function of the distribution of disease states within the population (Π), the patient's future referral and treatment decisions and the effort that specialists devote to treating the patient.

General practitioners

GPs are assumed to be risk-neutral. The Bernoulli utility function that represents their per-period preferences is simply their per-period profit. Assuming that they have a constant marginal cost of k per referral and no fixed costs, their per-period profits are

$$v(w, n_{j,1,t}) = n_{j,1,t}(w - k) = \sum_{i=1}^{I}(w - k)1_{i,j,t}.$$

The number of referrals a particular GP makes in period t is equal to the number of sick patients he has in that period ($n_{j,1,t}$). Note that the indicator variable ($1_{i,j,t}$) takes on the value one if patient i obtains a referral from GP j in period t and the value zero otherwise.

There are two sources of uncertainty that affect a GP's payoffs. First, there is the fact that $n_{j,1,t}$ is a random variable. We show later in this paper that this source of uncertainty disappears if the GP has an infinite patient pool, since the number of his patients who are sick in any particular period is almost surely infinite. The second source of uncertainty relates to the number of patients that will visit him in future periods. This may, in part, depend on the GP's success in motivating specialists to exert high effort when treating patients that are referred by him. If a GP makes his referral decision before observing the number of patients that visit him during period t, then his expected lifetime utility can be conveniently written as

$$V(\cdot) = V^B(\cdot) = E(n_{j,1,t} \,|n_{j,t})(w-k) + \delta Q = \pi n_{j,t} + \delta Q,$$

where Q is the GP's expected continuation payoff. However, if the GP does not make his referral decisions until after he has observed the number of patients that are seeking his referral services in the current period, then his lifetime expected utility becomes

$$V^A(\cdot) = n_{j,1,t}(w-k) + \delta Q.$$

As we show later in this paper, if a GP has an infinite patient pool, then $n_{j,1,t}$ converges almost surely to $\pi n_{j,t}$. As such, $V^A(\cdot)$ converges almost surely to $V^B(\cdot)$. Since we are interested in the case in which GPs have infinite patient pools, we will focus on $V^B(\cdot)$.

Medical Specialists

We will assume that all medical specialists have the same per-period preferences. These preferences are defined over the price they receive for providing treatment (r) and the effort they devote to that treatment (e). We will assume that these preferences may be represented by a quasi-linear Bernoulli utility function that is additively separable across patients. Each specialist's per-period per-patient preferences are given by

$$\widehat{z}(r,e) = r - C(e),$$

where treatment effort can either be high ($e = 1$) or low ($e = 0$). The cost of low effort, $C(0)$, will be normalised to zero, while the cost of high effort is $C(1) = C > 0$. Note that an implication of additive separability across patients is that the marginal disutility of effort is constant. It does not vary with the total amount of effort being exerted on all of the patients treated by a specialist in any given period. Let $n_{k,t}$ denote the number of patients a particular specialist has in period t and the variable $1_{i,k,t}$ indicate whether or not patient i was treated by this specialist k in period t. Since specialist per-period preferences are additively separable across patients, they may be represented by a per-period Bernoulli utility function of the form

$$z(r,e) = \sum_{i=1}^{I} \{r - C(e_{i,k,t})\} 1_{i,k,t} = n_{k,t}r - \sum_{i=1}^{I} C(e_{i,k,t}) 1_{i,k,t}.$$

As was the case with the GPs, the only uncertainty that affects medical specialists relates to the number of patients that will seek their treatment services in future periods. This may depend on a number of factors, including the specialist's current effort choices, the actual health outcomes following low effort treatment and the incidence of the disease in future periods. The specialist's expected lifetime utility can be conveniently written as

$$Z(\cdot) = n_{k,t}r - \sum_{i=1}^{I} C(e_{i,k,t}) 1_{i,k,t} + \delta Y,$$

where Y denotes the specialist's expected continuation payoff. This payoff will clearly depend on the number of patients that visit him in the future. While this is in part random, varying with Nature's decisions about disease incidence, it will also depend on the future decisions of patients and GPs.

2.3.3 The nature of equilibria

In this paper, we impose some sequential rationality restrictions on the set of acceptable equilibria for the supergame. We do this by solving the game by backwards induction. Since the market process in each period involves incomplete information that is not always revealed following its completion, there will be many non-singleton information sets in the supergame. As such, we would expect the set of equilibrium strategy profiles to depend on players' beliefs about the prior history of the game at each of their information sets. In an infinitely repeated game, these beliefs can be rather complicated.

In order to avoid the complicated beliefs that can arise in infinitely repeated games, we will make use of the competitive nature of our model of health care markets. In particular, we will solve the model in three stages. First, we will consider an infinitely repeated game between a representative patient and a representative specialist in the absence of GPs. This will provide a benchmark for the outcome if a patient chooses to self-refer. We will then consider an infinitely repeated game between a representative GP and a representative specialist. We will assume that the GP has a constant patient pool of infinite size. This will provide a benchmark for the outcome if a patient seeks a referral from a GP. Finally, we will consider a representative patient's choice between self-referral and seeking a referral from a GP.

2.4 The patient-specialist supergame

Our explanation for the structure of gated industries focuses on the role of intermediaries in ameliorating the market failure resulting from a static moral hazard problem. In order to pursue this line of reasoning, we need to understand the outcomes that result in such industries when intermediaries are not present.

These outcomes are analysed in this section.

Suppose that there are no GPs. In these circumstances, the only decisions that a patient makes are whether to seek treatment and, if so, which specialist to visit. The only decisions that a specialist needs to make involve the amount of effort to devote to treating each patient that visits him.

A desire to impose some credibility restrictions on the use of punishment threats is implicit in our decision to solve the entire supergame by backwards induction. We will maintain this approach within the patient-specialist supergame. In order to impose some degree of sequential rationality on the set of acceptable equilibria for the patient-specialist supergame, we will restrict our attention to perfect public equilibria.[10] Perfect public equilibria have two desirable properties, both of which simplify the process of finding sequentially rational Nash equilibria for a supergame with imperfect monitoring. The first property is belief independence. It is known that beliefs exist which will support a perfect public equilibrium as a perfect Bayesian equilibrium.[11]. As such, we know that any perfect public equilibrium is sequentially rational without needing to calculate the actual beliefs that make it so. The other desirable property of perfect public equilibria is that they are recursive, in the sense that, from any point in time, they will induce a perfect public equilibrium in every subsequent continuation game.

However, we would like to extend this backwards induction reasoning to the stage game itself. Recall that each specialist gets to make all of his effort choices after he observes which patients are seeking treatment from him in that period, as well as any continuation payoffs that the patients can credibly promise. As such, we will solve each specialist's problem first, conditional on

[10] The seminal papers on the perfect public equilibrium concept are Abreu et al ([2]) and Fudenberg et al ([39]). Useful discussions of the concept can be found in Fudenberg and Tirole ([40], chapter 5, sections 5 and 6) and Mailath and Samuelson ([66], chapter 7).

[11] See Fudenberg et al ([39], pp. 8-9). More recent work on belief-free equilibria can be found in Ely et al ([37]).

the patients' strategy choices. We will then solve each patient's problem under the assumption that the specialists will respond accordingly.

Specialists' strategy choices in the stage game simply amount to a choice of treatment effort for each patient that seeks treatment from them. We assumed earlier that the specialists' payoffs were additively separable across patients. Furthermore, patients cannot directly communicate treatment outcomes with each other in this model. As such, there is no direct gain for a specialist from linking his effort choices across patients. We will assume throughout this section that each specialist chooses the effort he will devote to treating each of his patients independently of the effort devoted to treating his other patients. When choosing the amount of effort to devote to treating a particular patient, the specialist will simply way up the expected lifetime utility of exerting high effort against that of exerting low effort. In each case, the expected lifetime utility will clearly depend on the expected continuation payoffs promised by the patient.

Patients' strategy choices consist of three components in this model. These components are a treatment decision, the choice of specialist in the event that treatment is chosen and a credible statement about their future treatment and specialist choices if they happen to get sick again. These future strategy choices can be represented by the choice of a continuation payoff for that specialist. This continuation payoff can vary with treatment outcomes.

2.4.1 The treatment choices of specialists

Since we wish to solve the specialist's problem first, suppose that a representative patient (i) who is sick has decided to seek treatment from some specialist (k). The patient will be able to motivate high effort from this specialist if and only if he can credibly promise continuation payoffs that will ensure that both the specialist's high effort incentive compatibility constraint and participation constraint are satisfied. Recall that the patient only observes the outcome of the treatment and not the effort devoted to treatment by the specialist. Furthermore, the patient only observes his own health outcomes and not those of other patients treated by the specialist. Similarly, the specialist only observes the patient's health outcomes when he treats the patient and not when the patient is treated by another specialist. Since we are restricting our attention to public strategies, the continuation payoffs can only be conditioned on the patient's history of health outcomes following treatment by this specialist. With prices fixed, the only punishment available to a patient is to dump the treating specialist.

This dumping strategy could be employed temporarily, with the patient refusing to visit that particular specialist at any time in the next T periods or the next T times he is sick. Alternatively, it could be employed permanently, with the patient refusing to ever seek treatment from that specialist again. The patient could choose to trigger the punishment only after a series of bad outcomes, or if the proportion of bad outcomes exceeds some threshold. Alternatively, the patient could trigger the punishment after only a single bad outcome. The most extreme punishment that could be chosen involves the patient permanently dumping the specialist if there is ever a bad outcome.[12] The extreme punishment strategy requires the specification of two continuation payoffs, one for histories in which the patient always has good outcomes following treatment ($V(1)$) and one for histories in which there is at least one bad outcome ($V(0)$). We will focus on this strategy in the analysis below.

Proposition 1 *The specialist will prefer to provide high effort treatment to a patient rather than low effort treatment if and only if the treatment price matches or exceeds some threshold price.*

[12] Recall that in this model, bad health outcomes following treatment can only occur if the specialist devotes low effort to that treatment.

Proof. Under the extreme punishment strategy employed by the patient, the specialist's high effort incentive compatibility constraint is:

$$r - C + \delta V(1) \geq r + \delta \left[\theta V(1) + (1-\theta)V(0)\right],$$

which simplifies to:

$$V(1) \geq V(0) + \frac{C}{\delta(1-\theta)}.$$

Since punishment involves dumping the specialist forever, we can set $V(0) = 0$. Furthermore, since the price of treatment is exogenous, the highest continuation payoff for a history of only good outcomes that a patient could credibly offer is a constant stream of the static payoff to high effort. This is:

$$V(1) = \sum_{t=0}^{\infty} \delta^t \pi(r - C) = \frac{\pi(r-C)}{(1-\delta)}.$$

Substituting these continuation payoffs into the high effort constraint yields:

$$\frac{\pi(r-C)}{(1-\delta)} \geq \frac{C}{\delta(1-\theta)}.$$

After some rearranging, this becomes:

$$r \geq \left[\frac{(1-\delta) + \pi\delta(1-\theta)}{\pi\delta(1-\theta)}\right] C,$$

or alternatively:

$$r \geq \left[1 + \frac{(1-\delta)}{\pi\delta(1-\theta)}\right] C.$$

We will call this inequality the threshold price inequality. ■

The threshold price referred to in proposition 1 is the lowest price at which a specialist will be willing to provide high effort treatment rather than low effort treatment. Specifically, in the absence of GPs, the threshold price is given by:

$$\widehat{r} = \left[1 + \frac{(1-\delta)}{\pi\delta(1-\theta)}\right] C.$$

Recall that $\pi \in (0, 1)$ and $\theta \in (0, 1)$. Furthermore, specialists are neither perfectly patient nor perfectly impatient, so that $\delta \in (0, 1)$. As such, the threshold price inequality in Proposition 1 implies the following result.

Proposition 2 *If high effort treatment is to be provided, then the price of such treatment must exceed the marginal cost of such treatment.*

This result is somewhat unusual for a competitive economy. It is generated by the asymmetric information that is present in the market for treatment services. The gap between the threshold treatment price and the disutility incurred by a specialist that provides high effort treatment is an information rent that must be paid in order to induce specialists to provide high effort treatment.

While the threshold price inequality in Proposition 1 guarantees that any specialist that provides treatment will prefer providing high effort treatment to low effort treatment, we still need to establish the circumstances under which a specialist would want to provide treatment of either variety. To simplify matters, we will assume that the specialist's reservation utility has been normalised to zero.

First, let us establish conditions under which the specialist will prefer to provide high effort treatment than provide no treatment whatsoever. If the specialist refuses to treat a patient, he will receive no surplus from that transaction. It is possible that the patient could punish such behaviour in a similar way to the punishment used for a bad health outcome from treatment. However, no such punishment is necessary to induce treatment in those cases where the patient can motivate high effort treatment from the specialist.

Proposition 3 *If the high effort incentive compatibility condition holds, then the specialist will prefer to provide high effort treatment to the patient rather than not treating the patient at all.*

Proof. We know from Proposition 2 that if the high effort incentive compatibility constraint is satisfied, then $r > C$. This is sufficient to ensure that the specialist would receive positive surplus if he provides high effort treatment to the patient. Since the specialist receives no surplus if he refuses to treat the patient, he will prefer to provide high effort treatment to the patient rather than not treat the patient at all. ■

Note that the result in Proposition 3 holds, even if no dynamic punishment for non-treatment is used by the patient. If patients are able to motivate high effort from the specialist, they can automatically ensure participation.

Suppose instead that the high effort incentive compatibility constraint does not hold. In this case, the specialist will only provide low effort treatment, if any treatment is provided at all.

Proposition 4 *If the high effort incentive compatibility constraint does not hold, then the specialist will (weakly) prefer to provide low effort treatment to the patient over not treating the patient at all if the prevailing treatment price is non-negative.*

Proof. If the specialist provides low effort treatment, then he only incurs the disutility associated with low effort treatment. Thus his cost of treatment is $C(0) = 0$. As such, any non-negative price for treatment will be sufficient to induce the specialist to offer treatment, even if the patient does not employ any dynamic punishments for non-treatment. ■

Note that if the treatment price is positive, then the specialist will earn a positive surplus from the transaction. Even in the absence of dynamic punishments for non-treatment, this still exceeds the surplus from non-treatment, which is zero. If the price is zero, then in the absence of dynamic punishments for non-treatment, the specialist will be indifferent between providing low effort treatment to the patient and not treating the patient at all. We will adopt the standard convention and assume that when the specialist is indifferent between providing low effort treatment and no treatment, the specialist chooses to provide low effort treatment.

The results in Proposition 3 and Proposition 4 ensure that motivating specialists to provide treatment is not a problem in this economy. The only question is whether they will provide high effort treatment or low effort treatment. The conditions under which high effort treatment will be provided are given by the threshold price inequality in Proposition 1. If this high effort incentive compatibility condition does not hold, then low effort treatment will be provided.

2.4.2 The treatment choices of patients

Patient's preferences depend on both their health state and their expenditure on health care. A sick patient will only seek treatment if the expected benefits in terms of a higher probability of good health exceed the cost of the treatment. A patient who is not sick will not seek treatment, since doing so will involve a cost but yield no benefit. As such, we will focus on the treatment choices of a sick patients. Since there are no GPs present in this hypothetical economy, the patient cannot seek a referral. As such, if the patient chooses to seek treatment, the only expenditure incurred will be the treatment price, so that $p = r$. Prior to seeking treatment in any given period, a sick patient does not know the amount of effort that will be exerted by the treating specialist. As such, his expected utility from treatment is:

$$U(h,p) = \mu B - r + \delta M,$$

where M is the patients expected continuation payoff if he seeks treatment in the current period and $\mu = \lambda + \theta(1-\lambda)$ is the patient's belief that he will be cured following treatment in the current period.

In principle, we could allow the continuation payoff to vary with the decision to seek treatment for disease k and the health state following any such treatment in every period up until and including the current period. The reason for this is that these will be observed by both the representative patient and the representative specialist. However, given the competitive nature of this model, we will assume that specialists do not condition their future strategy choices on the history of treatment choices or the health outcomes of their patients in the current period. As such, from a patient's point of view, the continuation payoff does not vary with the public history of either the treatment choices for disease k or the public history of health outcomes following any such treatment. Hence we can set all of the patient's continuation payoffs in the current period equal to M. Given this, the patient's expected utility if he does not seek treatment is:

$$U^0 = \delta M.$$

Proposition 5 *If the treatment price is not too high, a sick patient will seek treatment.*

Proof. A sick patient will seek treatment if and only if the following individual rationality constraint is satisfied:

$$\mu B - r + \delta M \geq \delta M.$$

This constraint simplifies to the following restriction on the treatment price:

$$r \leq \mu B.$$

Thus, so long as the treatment price does not exceed the expected benefit to the patient from treatment, he will seek treatment. ■

A patient who is using the extreme dumping strategy outlined previously will know whether or not the treatment price is at least as large as the high effort threshold price. As such, the patient will know whether or not he will receive high effort treatment. This allows us to be more specific about the patient's decision to seek treatment.

Proposition 6 *If the treatment price matches or exceeds the threshold treatment price, a sick patient will seek treatment whenever* $r \leq B$*. If the treatment price is less than the threshold treatment price, a sick patient will seek treatment whenever* $r \leq \theta B$*.*

Proof. If the treatment price matches or exceeds the threshold price, then the patient knows that he will receive high effort treatment. As such, $\lambda = 1$ and hence $\mu = 1$. Thus the maximum treatment price that the patient will be willing to pay in this case is $r = B$. If the treatment price is less than the threshold price, then the patient knows that he will receive low effort treatment. As such, $\lambda = 0$ and hence $\mu = \theta$. Thus the maximum treatment price that the patient will be willing to pay in this case is $r = \theta B$. ■

Clearly, the extreme dumping strategy is designed to induce the specialist to provide high effort treatment. Assuming that a patient uses the extreme dumping strategy, we have characterised the range of prices for he will seek and receive high effort treatment. We have also characterised the range of prices for which he will seek and receive low effort treatment. However, we have not yet established that a sick patient would prefer high effort treatment to low effort treatment.

Proposition 7 *A sick patient will always prefer to receive high effort treatment rather than low effort treatment for any given treatment price.*

Proof. The payoff to a sick patient who receives high effort treatment is $B-r+\delta M$. The payoff to a sick patient who receives low effort treatment is $\theta B-r+\delta M$. The patient will prefer high effort treatment over low effort treatment if and only if:

$$B - r + \delta M \geq \theta B - r + \delta M.$$

This expression simplifies to:

$$(1-\theta)\,B \geq 0.$$

Since $\theta \in (0,1)$ and $B > 0$, this inequality is always satisfied. As such, a sick patient will prefer high effort treatment to low effort treatment for any given treatment price. ■

We now know that a patient will prefer high effort treatment to low effort treatment and that he can motivate a specialist to provide high effort treatment if the treatment price is sufficiently high.

2.4.3 Long-run and short-run relationships

If a patient is able to motivate high effort treatment from the specialist, then he is said to have a long-run relationship with that specialist. If a patient is not able to motivate high effort treatment from a specialist, then he is said to have a short-run relationship with that specialist. We have already found conditions on the prevailing treatment price that will allow us to characterise the relationship between a patient and a specialist as either long-run or short-run. However, it is perhaps more intuitive to define a short-run relationship between a patient and a specialist in terms of the probability that a patient will need the services of a specialist in any given period. After all, if that probability is sufficiently low, a patient that is being treated in the current period will be unlikely to require treatment for the foreseeable future. Given that the specialist is not perfectly patient ($\delta < 1$), he is likely to ignore any impact on this patient's future demand for his services when choosing his current effort level. Recall that threshold treatment price for ensuring high effort treatment was a function of the probability that the patient will get sick in any given period. As such, we can rearrange the high effort incentive compatibility condition to provide a restriction on the probability that a patient gets sick in any given period.

Proposition 8 *A patient has a short-run relationship with a specialist if and only if at least one of the following three conditions hold: (a)* $(r - C) \leq 0$*, (b)* $\pi < \widehat{\pi}$ *and (c)* $\widehat{\pi} > 1$*.*

Proof. Recall that a patient can motivate high effort treatment from a specialist if and only if:

$$r \geq \left[1 + \frac{(1-\delta)}{\pi\delta(1-\theta)}\right] C.$$

Since this requires that $(r - C) > 0$, we know that the patient will receive low effort treatment if $(r - C) \leq 0$. We can rearrange the threshold price inequality to obtain:

$$\pi \geq \widehat{\pi} = \frac{(1-\delta)C}{\delta(1-\theta)(r-C)}.$$

Thus a patient will have a short-run relationship with a specialist if $\pi < \widehat{\pi}$. Finally, note that $\pi \in [0, 1]$ since it is a probability. As such, if $\widehat{\pi} > 1$, then the patient will have a short-run relationship with the specialist. ■

If none of the conditions in Proposition 8 hold, then the patient will have a long-run relationship with the specialist.

Proposition 9 *The patient will have a long-run relationship with the specialist if and only if all of the following conditions hold: (a)* $(r - C) > 0$*, (b)* $\pi \geq \widehat{\pi}$ *and (c)* $\widehat{\pi} \leq 1$.

2.5 The GP-specialist supergame

Suppose that patients only have short-run relationships with specialists. They might be willing to seek a referral from a GP if they thought that this would result in high effort treatment and obtaining the referral was not too costly. In this section, we examine the circumstances under which a GP will be able to motivate a specialist to provide high effort treatment to all of the patients that are referred to the specialist by him. In order to incorporate the idea that each GP has a large patient pool that is stable in size, we will ultimately assume that each GP has an infinite patient pool. This assumption is required to ensure that it will be rational for specialists to hold static expectations with respect to the size of GP patient pools. A specialist has static expectations about a GP's patient pool if, in any given period, he believes that the number of patients utilising the GP's services will remain at its level in that period forever. We will begin the analysis by assuming that the representative GP has a finite patient pool of size n. The infinite patient pool assumption will be implemented by taking limits as $n \to \infty$.

2.5.1 Optimal deviation by a specialist

Consider a representative GP, j, who currently has a patient pool of size n. Suppose that $n_{j,1}$ of these patients are sick in the current period and that the GP j chooses specialist k to treat all of these patients. Can the GP motivate the specialist to exert high effort whenever the specialist is treating patients referred by him? Before considering this question, we will need to make some simplifying assumptions about the nature of competition. First, we continue to assume that all agents are price takers and that prices are set exogenously. In addition to this, we will assume that specialists have static expectations with respect to the size of GP patient pools. In any given period, they believe that the number of patients utilising a GP's services will remain at its current level forever. Justifications for this assumption are provided later in this paper, when the price formation process is considered.

Suppose that a specialist decides to deviate and shirk in his treatment of at least one of GP j's patients. What is the specialist's optimal deviation? This amounts to determining how many of the $n_{j,1}$ patients should receive low effort treatment.

Proposition 10 *If a specialist chooses to shirk when treating any patient referred to him by a particular GP in a particular period, then he will shirk when treating every patient referred to him by that GP in that period.*

Proof. If $n_{j,1} = 1$, this question is easy to answer. The only possible deviation from high effort treatment for all of GP j's patients is to shirk for that lone patient. When $n_{j,1} > 1$, the specialist could choose to provide low effort treatment for all of these patients or just for some subset of them. From the specialist's point of view, all of the patients referred by a particular GP in any given period are identical. Thus we need only consider the number of these patients that receive low effort treatment and not their individual identities. Let $m_{j,1}$ denote the number of patients referred to the specialist by GP j in the current period that receive low effort treatment. If the GP employs an all or nothing punishment strategy, then only two continuation payoffs need to be specified. These are the payoff to only good health outcomes, $V(1)$, and the payoff if there are any bad outcomes, $V(0)$.

We will assume that the GP follows up on the treatment outcomes for all of the patients he refers at the end of each period. Thus the GP can condition the specialist's continuation payoffs on whether or not a bad outcome occurs for any of the patients he referred to the specialist in the current period or in any past period. Given this, the payoff to the specialist from providing low effort treatment to $m_{j,1}$ of the $n_{j,1}$ patients referred by GP j is

$$\widehat{U}_s(m_{j,1}; n_{j,1}) = n_{j,1}r - (n_{j,1} - m_{j,1})\, C + \delta \left[\theta^{m(j,1)} V(1) + \left(1 - \theta^{m(j,1)}\right) V(0)\right].$$

If the GP's punishment for a bad outcome is to sack the specialist, so that $V(0) = 0$, this becomes

$$\widehat{U}_s(m_{j,1}; n_{j,1}) = n_{j,1}r - (n_{j,1} - m_{j,1})\, C + \delta\theta^{m(j,1)} V(1).$$

Differentiating this with respect to the number of patients for which the specialist shirks ($m_{j,1}$), we obtain

$$\frac{\partial \widehat{U}_s(m_{j,1}; n_{j,1})}{\partial m_{j,1}} = C + \delta\theta^{m(j,1)} \ln(\theta) V(1).$$

Since $\theta \in (0, 1)$, so that $\ln(\theta) < 0$, the sign of this derivative is ambiguous. We could assume that the derivative is always positive, but that would place strong restrictions on the size of the disutility of effort. Notice that, since $\ln(\theta) > -\infty$

because $\theta > 0$, we have

$$\lim_{m(j,1)\to\infty} \delta\theta^{m(j,1)} \ln(\theta)V(1) = \delta(0)\ln(\theta)V(1) = 0.$$

Hence, for sufficiently large $m_{j,1}$, the derivative will be positive. Indeed, the derivative is monotonically increasing in $m_{j,1}$, since

$$\frac{\partial^2 \widehat{U}_s(m_{j,1}; n_{j,1})}{\partial m_{j,1}^2} = \delta\theta^{m(j,1)} \left[\ln(\theta)\right]^2 V(1) > 0 \text{ for all } m_{j,1}.$$

If the first derivative of $\widehat{U}_s(m_{j,1}; n_{j,1})$ with respect to $m_{j,1}$ is positive for all $m_{j,1}$, then the unique optimal deviation is for the specialist to shirk for all of GP j's patients. If the derivative is negative for sufficiently low $m_{j,1}$, then there are two possibilities for the optimal deviation. One possibility is that the specialist will not want to shirk at all, so that $m_{j,1} = 0$, in which case there is no deviation. However, if the specialist is going to shirk for any of GP j's patients, he will choose to shirk for all of them. The optimal deviation is thus $m_{j,1} = n_{j,1}$. ■

The intuition behind this result is clear. The marginal benefit of shirking is simply the avoided cost of effort for the additional patient (C), which is constant. It does not change as the number of patients referred by GP j that receive low effort treatment increases. However, the additional probability of being detected (and hence the expected marginal cost of shirking) falls as the number of patients referred by GP j that receive low effort treatment increases. As such, if the specialist chooses to shirk when treating any of GP j's patients in any given period, he will shirk when treating all of GP j's patients in that period.

We are now in a position to derive the maximum payoff that a specialist will receive if he shirks for any of GP j's patients when GP j is employing the extreme punishment strategy outlined earlier.

Proposition 11 *If a GP employs the extreme punishment strategy, the maximum payoff to a specialist who fails to provide high effort treatment for all of the patients referred by that GP is*

$$\widehat{U}_s(deviate) = n_{j,1}r + \delta\theta^{n(j,1)}V(1).$$

Proof. We have already shown that the optimal deviation for a specialist involves setting $m_{j,1} = n_{j,1}$. Substituting this into the expression for the specialists payoff yields

$$\widehat{U}_s(deviate) = \widehat{U}_s(n_{j,1}; n_{j,1}) = n_{j,1}r + \delta\theta^{n(j,1)}V(1).$$

■

A GP will be able to motivate high effort treatment from a specialist for all of his patients if and only if the payoff to specialist from providing only high effort treatment exceeds both the payoff to his optimal deviation and the payoff to refusing

to treat the patients. The conditions under which a specialist would prefer to provide high effort treatment to all of the patients referred from GP j are derived in the next subsection of this paper. The conditions under which the specialist will choose to treat all of GP j's patients are considered in the following subsection of this paper.

2.5.2 Incentive compatibility constraints for specialists

A specialist will provide high effort treatment to all of the patients that are referred to him by a particular GP only if the payoff to doing so exceeds the payoff that he would receive from shirking for at least some of these patients. This requires that the treatment price be sufficiently high.

Proposition 12 *A specialist will weakly prefer to provide high effort treatment to all of the patients that are referred to him by a particular GP in a given period over the provision of low effort treatment to one or more of these patients if the treatment price exceed some threshold price. The threshold price is given by*

$$\widehat{r}(n_{j,1}, n_j) = \left[1 + \frac{n_{j,1}(1-\delta)}{\pi_j n_j \delta \left(1 - \theta^{n(j,1)}\right)}\right] C.$$

Proof. The payoff to the specialist from providing high effort treatment for all of GP j's patients is

$$\widehat{U}_s(no\ deviation) = n_{j,1}(r - C) + \delta V(1).$$

Thus the specialist will choose to provide high effort treatment to every patient referred by GP j only if the payoff from doing so matches or exceeds the largest possible payoff from not doing so. This requires that

$$n_{j,1}(r - C) + \delta V(1) \geq n_{j,1} r + \delta \theta^{n(j,1)} V(1),$$

which can be rearranged to obtain

$$V(1) \geq \frac{n_{j,1} C}{\delta \left(1 - \theta^{n(j,1)}\right)}.$$

■

In determining the equilibrium continuation payoff, $V(1)$, we need to remember that the specialists have static beliefs about the size of each GP's patient pool, n_j.[13] Given this, the highest continuation payoff for a history of only good

[13] Strictly speaking, this only makes sense if GPs have infinite patient pools. All that specialists observe is the number of patients that are referred to them by a particular GP in any given period ($n_{j,1}$). As such, they need to infer the size of the GP's patient pool (n_j) on the basis of this information. Since $n_{j,1} \sim bin(\pi, n_j)$, the specialist will view n_j as a non-

outcomes that a GP can credibly offer is a constant stream of the expected static payoff to high effort. This is

$$V(1) = \sum_{t=0}^{\infty} \delta^t \pi n_j (r - C) = \frac{\pi n_j (r - C)}{(1 - \delta)}.$$

Note that $(r - C)$ is the net payoff per patient when the specialist exerts high effort, while πn_j is the expected number of GP j's patients that will be sick in any given period. Substituting this into the high effort ICC we obtain

$$\frac{\pi n_j (r - C)}{(1 - \delta)} \geq \frac{n_{j,1} C}{\delta \left(1 - \theta^{n(j,1)}\right)},$$

which can be rearranged to yield

$$r \geq \left[1 + \frac{n_{j,1} (1 - \delta)}{\pi n_j \delta \left(1 - \theta^{n(j,1)}\right)}\right] C.$$

If the prevailing treatment price does not fall below this threshold, then the GP will be able to assure his patients that they will provided with high effort treatment by this specialist in the current period.

We will denote the threshold price, below which a GP cannot ensure high effort treatment for all of his patients, by $\widehat{r}(n_j, n_{j,1})$. Note that when GPs have finite patient pools, this threshold price is a random variable. The reason for this is that the threshold price is a function of the number members of a GP's patient pool who are sick in a particular period. This means that a patient cannot be sure that any particular GP will be able to motivate high effort treatment on the part of a specialist, even if he knows both the prevailing treatment price and the size of each GPs patient pool. This situation can be avoided if GPs have

degenerate random variable if the GP has a finite patient pool. However, if the GP has an infinite patient pool, then $n_{j,1}$ is almost surely infinite. This greatly simplifies the statistical inference problem facing the specialist. If a specialist receives an infinite number of referrals from a particular GP, then he knows that the GP has an infinite patient pool.

infinite patient pools.

Proposition 13 *If each GP has an infinite patient pool, then the threshold treatment price is almost surely* $\frac{C}{\delta}$.

Proof. Consider a GP who has a patient pool of size n_j. Suppose that $n_{j,1}$ of these patients are sick in a particular period. The threshold treatment price for such a GP will be

$$\widehat{r}(n_j, n_{j,1}) = \left[1 + \frac{n_{j,1}\,(1-\delta)}{\pi n_j \delta \left(1 - \theta^{n(j,1)}\right)}\right] C = \left[1 + \frac{\alpha_j\,(1-\delta)}{\pi \delta\,(1-\beta_j)}\right] C,$$

where $\alpha_j = \frac{n_{j,1}}{n_j}$ is the proportion of the GPs patients who are sick in that period and $\beta_j = \theta^{n(j,1)}$ is the probability that a specialist who shirks when treating all of these patients does not get caught. Determining what happens to the threshold price as the size of a GP's patient pool approaches infinity requires us to determine what happens to $n_{j,1}$ as $n_j \to \infty$. This is not straightforward, as the relationship between $n_{j,1}$ and n_j is stochastic. Indeed, $n_{j,1}$ can be viewed as the number of negative outcomes in a random sample of n_j Bernoulli trials, where the probability of a negative outcome on any given trial is π. As such, $n_{j,1}$ is a binomially distributed random variable, with parameters n_j and π. The specific number of patients that are referred to a specialist by a GP in any given period is simply a particular realisation of this underlying random variable. Unfortunately, it is this actual realisation that enters a specialist's high effort incentive compatibility constraint and hence the threshold price. In finite samples, any particular realisation of $n_{j,1}$ could occur with positive probability. However, in an infinite sample, we can use limiting arguments to show that the relative proportion of negative outcomes (α_j) is almost surely equal to the probability of a negative outcome in a single trial (π). This in turn allows us to show that each GP almost surely has an infinite number of sick patients in each period. Furthermore, the probability that any specialist who shirks when treating all of these patients is not caught is almost surely equal to zero. The combination of these limiting results allows us to show that the threshold price for any GP who has an infinite patient pool is almost surely a constant.

First, we need to show that α_j almost surely converges to π as n_j approaches infinity. Recall that

$$\alpha_j = \frac{n_{j,1}}{n_j} = \frac{1}{n_j} \sum_{i(j)=1}^{n(j)} 1_{i(j),1},$$

where $1_{i(j),1}$ is an indicator variable that takes on the value one if a particular member of GP j's patient pool, patient $i(j)$, has the disease in the current period and zero otherwise. The summation is over GP j's entire patient pool for the current period. Not that each of these indicator variables is a Bernoulli random variable that takes on the value one with probability π and zero otherwise. As such, $\left\{1_{i(j),1}\right\}_{i(j)=1}^{n(j)}$ is a sequence of independent and identically distributed Bernoulli random variables. Furthermore, note that

$$E\left(1_{i(j),1}\right) = \pi(1) + (1-\pi)(0) = \pi \text{ for all } i_j \in \{1, 2, \cdots, n_j\}.$$

Thus, from the strong law of large numbers[14], we know that

$$\Pr\left(\lim_{n(j)\to\infty} \alpha_j = \pi\right) = 1.$$

[14] See Billingsley ([16], pp. 85-86) for a discussion of the strong law of large numbers. Note that when GP patient pools are finite, the number of members in a GP's patient pool is an integer. As such, when we take the limit as this number approaches infinity, we are restricting our attention to the set of natural numbers. In effect, there is a one-to-one correspondence between each member of a GP's patient pool and each element of the set of natural numbers when that GP has an infinite patient pool. As such, each GP has a countable number of patients. This ensures that the standard version of the strong law of large numbers applies in the model considered in this paper. If we had assumed that each GP had a continuum of patients instead of a countably infinite number of patients, we would have needed to use the techniques mentioned in Judd ([57]). The reason for this is that each GP would have had an uncountable number of patients in that case.

Hence we can conclude that the relative proportion of sick patients in any given period for a particular GP (α_j) is almost surely equal to the probability that any individual patient is sick in any given period (π) if the GP has an infinite patient pool. Now we want to show that the probability that any specialist who shirks when treating all of a GPs patients in any given period is not caught is almost surely equal to zero when the GP has an infinite patient pool. Since $\beta_j = \theta^{n(j,1)}$ and $\theta \in (0,1)$, this will be clearly be the case if $n_{j,1}$ approaches infinity as n_j approaches infinity. Note that

$$n_{j,1} = \left(\frac{n_{j,1}}{n_j}\right) n_j = \alpha_j n_j.$$

Furthermore,

$$\lim_{n(j)\to\infty} n_j = \infty.$$

Thus we can conclude that

$$n_{j,1} \xrightarrow{a.s.} (\pi)(\infty) = \infty.$$

Hence we know that we know that the probability that any specialist who shirks when treating all of a GPs patients in any given period is not caught is almost surely equal to zero when the GP has an infinite patient pool.

We are now in a position to look at what happens to the threshold price, $\widehat{r}(n_j, n_{j,1})$, as the size of GP's patient pools get very large. Note that $\widehat{r}(n_j, n_{j,1})$ is a continuous function of α_j and β_j as long as $\beta_j \neq 1$. Furthermore, since $\theta \in (0,1)$ ensures that $\beta_j \in [0,1)$, we do not need to worry about the potential discontinuity at $\beta_j = 1$. This means that

$$\widehat{r}(n_j, n_{j,1}) \xrightarrow{a.s.} \left[1 + \frac{\pi(1-\delta)}{\pi\delta(1-0)}\right] C = \left[\frac{\delta + 1 - \delta}{\delta}\right] C = \frac{C}{\delta}.$$

Thus we have established that the threshold treatment price for any GP with an infinite patient pool is almost surely $\frac{C}{\delta}$. ■

2.5.3 Participation constraints for specialists

While we have established the conditions under which a specialist will prefer providing high effort treatment to low effort treatment, we still need to establish that the specialist would prefer providing high effort treatment for all of the patients referred by a particular GP to not providing some or all of them with any treatment. If the specialist refuses to treat any of a GP's referrals, he will receive no surplus from that transaction. It is possible that the GP could punish such behaviour by refusing to refer any future patients to that specialist. However, as in the case without GPs, no such punishment is necessary to induce treatment in those cases where the GP can motivate high effort treatment from the specialist.

Proposition 14 *If the high effort incentive compatibility constraint is satisfied for a specialist with respect to a particular GP, then the specialist will prefer to provide high effort treatment to any patient referred by that GP in a given period, rather than not treat the patient at all.*

Proof. Since specialists are not perfectly patient ($\delta \in (0, 1)$), the threshold price must exceed the cost of high effort treatment. Thus we must have $r \geq \widehat{r}(n_j, n_{j,1}) > C$. This is sufficient to ensure that the specialist would receive positive surplus if he provides high effort treatment to the patient. Since the specialist will receive zero surplus from any patient he refuses to treat, he will prefer to provide high effort treatment rather than no treatment whatsoever. ■

Thus the specialist will prefer to provide high effort treatment to the patient than not treat the patient at all, even if no dynamic punishment for non-treatment is used by the referring GP. If a GP is able to motivate high effort from the specialist, he can automatically ensure participation.

Suppose instead that the high effort incentive compatibility constraint does not hold. In this case, the specialist will only provide low effort treatment to any patient referred by the GP in that period, if any treatment is provided at all.

Proposition 15 *If the high effort incentive compatibility constraint is not satisfied for a specialist with respect to a particular GP, then the specialist will weakly prefer to provide low effort treatment to any patient that is referred to him by that GP in a given period, rather than not treat the patient at all.*

Proof. If the specialist only provides low effort treatment for each of the patients referred by the GP, then he only incurs the disutility associated with low effort treatment, $C(0) = 0$, for each of those patients. As such, any non-negative price for treatment will be sufficient to induce the specialist to offer at least low effort treatment, even if the GP does not employ any dynamic punishments for non-treatment. ■

2.5.4 Participation constraints for GPs

GP's will be willing to provide referral services if and only if the discounted present value of their expected revenues exceeds that of their expected costs. Like all of the other players in this economy, GPs are price takers. As such, the only way their future revenue can be affected is by patients choosing not to utilise their referral services. Since prices and costs are exogenously fixed and constant across time in this economy, patients cannot induce specialists to provide treatment at a price below cost now in return for their future custom at above cost prices. Thus, GPs will provide their referral services if and only if the price per referral is at least as high as the cost per referral ($w \geq k$).

Proposition 16 *GPs will offer referral services if and only if the referral price exceeds the marginal cot of a referral.*

Proof. Recall that there are no fixed costs associated with providing referral services in this economy. Furthermore, the variable costs are constant. As such, the marginal cost of a referral equals the average cost per referral. Given this, the proposition follows from the above arguments. ■

2.6 Industry structure with exogenous prices

The structure of the health care industry will be jointly determined by the decisions of patients, general practitioners and specialists. We have characterised the conditions under which patients can motivate high effort treatment from specialists by themselves and the conditions under which GPs can motivate high effort treatment from specialists for all of their patients. We have also analysed the conditions under which GPs will be willing to offer their referral services and specialists will be willing to offer their treatment services. Finally, we have analysed the conditions under which a patient will demand treatment services alone. All that remains is for us to determine the circumstances under which a patient will prefer to seek both treatment and a referral over both treatment alone and no treatment whatsoever. We will then be in a position to describe how the structure of the health care industry will vary with both the treatment price and the referral price.

2.6.1 The referral choices of patients

In order to determine the circumstances under which a patient will seek a referral, we need to compare the payoff that a patient gets from obtaining a referral and treatment with both the payoff that the patient would get if he sought treatment alone and the payoff he would get without treatment. We can ignore the possibility that a patient will seek a referral alone because it would not improve his expected health status but it would use resources that could otherwise be spent on consumption. The treatment outcomes facing a patient who chooses not to seek a referral will be the same as those in the absence of GPs. Recall that these treatment outcomes varied with the prevailing treatment price as follows:

$$\textit{treatment outcome} = \begin{cases} \text{high effort treatment}, & \text{if } r \in [\widehat{r}_1, B]\,; \\ \\ \text{low effort treatment}, & \text{if } r \in [0, \min\{\widehat{r}_1, \theta B\})\,; \\ \\ \text{no treatment}, & \text{otherwise}; \end{cases}$$

where

$$\widehat{r}_1 = \left[1 + \frac{(1-\delta)}{\pi\delta(1-\theta)}\right] C.$$

The patient's continuation payoff is not affected by the current period outcome. As such, we can focus on the current period payoffs facing the patient. In the absence of a referral, these are

$$EU(h, r) = \begin{cases} B - r & \text{if } r \in [\widehat{r}_1, B]\,; \\ \\ \theta B - r & \text{if } r \in [0, \min\{\widehat{r}_1, \theta B\})\,; \\ \\ 0 & \text{otherwise.} \end{cases}$$

Now suppose that a patient seeks a referral. The treatment outcomes for a patient who has a referral are

$$treatment\ outcome = \begin{cases} \text{high effort treatment,} & \text{if } r \in \left[\frac{C}{\delta}, B\right]; \\ \\ \text{low effort treatment,} & \text{if } r \in \left[0, \min\left\{\frac{C}{\delta}, \theta B\right\}\right); \\ \\ \text{no treatment,} & \text{otherwise.} \end{cases}$$

The patient's payoffs if he seeks a referral are

$$EU(h, r+w) = \begin{cases} B - r - w & \text{if } r \in \left[\frac{C}{\delta}, B\right]; \\ \\ \theta B - r - w & \text{if } r \in \left[0, \min\left\{\frac{C}{\delta}, \theta B\right\}\right); \\ \\ 0 & \text{otherwise.} \end{cases}$$

Proposition 17 *A necessary condition for a patient to seek a referral is that* $r \in \left[\frac{C}{\delta}, \widehat{r}_1\right)$.

Proof. Clearly, if $r \in [\widehat{r}_1, B]$, then the patient will choose not to seek a referral if $w > 0$. When $w = o$, the patient will be indifferent between seeking a referral and self-referring. In these circumstances, we will assume that the patient self-refers. As such, whenever, $r \in [\widehat{r}_1, B]$, patients will not seek referrals. Furthermore, if $r \in \left[0, \frac{C}{\delta}\right)$, then patients who seek treatment will receive low effort treatment regardless of whether or not they have a referral. As such, these patients will not seek a referral either. Thus, a necessary condition for patients to seek a referral is that $r \in \left[\frac{C}{\delta}, \widehat{r}_1\right)$. ■

While this is a necessary condition for a patient to seek a referral, it is not a sufficient condition. When treatment prices satisfy $r \in \left[\frac{C}{\delta}, \widehat{r}_1\right)$, a patient will receive high effort treatment if he obtains a referral and low effort treatment if he self-refers. As such, his expected health benefits will be higher if he obtains a referral. However, his treatment costs will also be higher unless treatment is free. As such, a patient will seek a referral only if the additional expected benefits from receiving high effort treatment match or exceed the cost of a referral.

Proposition 18 *A patient will seek a referral if and only if both* $r \in \left[\frac{C}{\delta}, \widehat{r}_1\right)$ *and* $w \leq (1-\theta)\,B$.

Proof. we have already established that a patient will not seek a referral unless $r \in \left[\frac{C}{\delta}, \widehat{r}_1\right)$. Even if this condition is satisfied, the payoff to obtaining a referral must be at least as high as the payoff to self-referring if the patient is to seek a referral. This requires that

$$B - r - w \geq \theta B - r,$$

which can be rearranged to yield

$$w \leq (1-\theta)\,B.$$

■

2.6.2 The equilibrium industry structure

We have established the circumstances under which patients will seek a referral and treatment, seek treatment alone and seek neither treatment nor referral. We have also established the conditions under which GPs will offer their referral services and specialists will offer their treatment services. The market outcome will vary with the prevailing treatment and referral prices. The relationship between market outcomes and prices is summarised in Table 2.1.

Table 2.1: Market outcomes

Circumstances	**Market Outcomes**
$r \in \left[\frac{C}{\delta}, \widehat{r}_1\right), \, w \leq (1-\theta)\, B, \, r + w \leq B$	High effort treatment, referral
$r \in [\widehat{r}_1, B]$	High effort treatment, no referral
$r \in \left[0, \min\left\{\frac{C}{\delta}, \theta B\right\}\right)$	Low effort treatment, no referral
$r \in \left[\frac{C}{\delta}, \min\left\{\theta B, \widehat{r}_1\right\}\right), \, w < k$	Low effort treatment, no referral
Otherwise	No treatment, no referral

These market outcomes can be illustrated in (r, w)-space. A variety of possible outcomes are illustrated in Figures 2.1 to 2.3. Note that the presence of GPs allows for the existence of a region in (r, w)-space in which patients will choose to seek a referral. This will result in them getting high effort treatment where, in most cases, they would not do so otherwise. This provides the foundation for a demand driven explanation for the existence of GPs. When treatment and referral prices fall in this region, patients will prefer to have the option of seeking a referral from a GP. The reason for this is that GPs are able to motivate high effort treatment from specialists when prices fall in this region, while patients cannot do so. Furthermore, the additional health benefits that patients expect to receive from high effort treatment exceed the additional cost of seeking a referral when prices fall in this region.

In general, specialists do not like the presence of GPs. The reason for this is that they need to provide high effort treatment, which involves a higher disutility of effort for them, but they do not receive any additional remuneration. However, there are some circumstances in which both patients and specialists prefer to have GPs present. These situations involve treatment prices that satisfy $r \in \left(\frac{C}{\delta}, \theta B\right]$, where this interval is non-empty. In the absence of GPs, patients would not seek treatment and specialists would earn no profits. If GPs are present, then patients will seek both a referral and treatment. Specialist will provide high effort treatment and earn positive profits. If $k \leq w \leq \min\{(1-\theta)B, B-r\}$, then patients, GPs and specialists will all weakly prefer the presence of GPs to their absence in such circumstances. As such, there are some cases in which a gated industry structure weakly Pareto dominates an ungated industry structure when prices are exogenous. Circumstances such as these occur for some treatment prices in Figure 2.2.

2.7 Conclusion

We have used reputation effects to explain the organisation of many professional service industries, including the medical and legal professions. The main focus has been on explaining the existence of gatekeeping intermediaries who refer consumers to one of many ultimate producers. Examples of such intermediaries include general practitioners in the health care industry and solicitors in the legal industry. The explanation for the existence of such intermediaries that is provided in this paper focuses on their role as a reputation monitor. The GPs keep track of the treatment outcomes for each patient they refer to a particular specialist. GPs have large patient pools because they provide referral services for many different types of disease. As such, they will observe many more treatment outcomes with a particular specialist than any individual patient will observe. Furthermore, the fact that a GP has a large patient pool also means that if he discovers evidence of shirking on the part of the specialist, he can

Figure 2.1: Market outcomes when $\theta B < \frac{C}{\delta} < \widehat{r}_1 < B$.

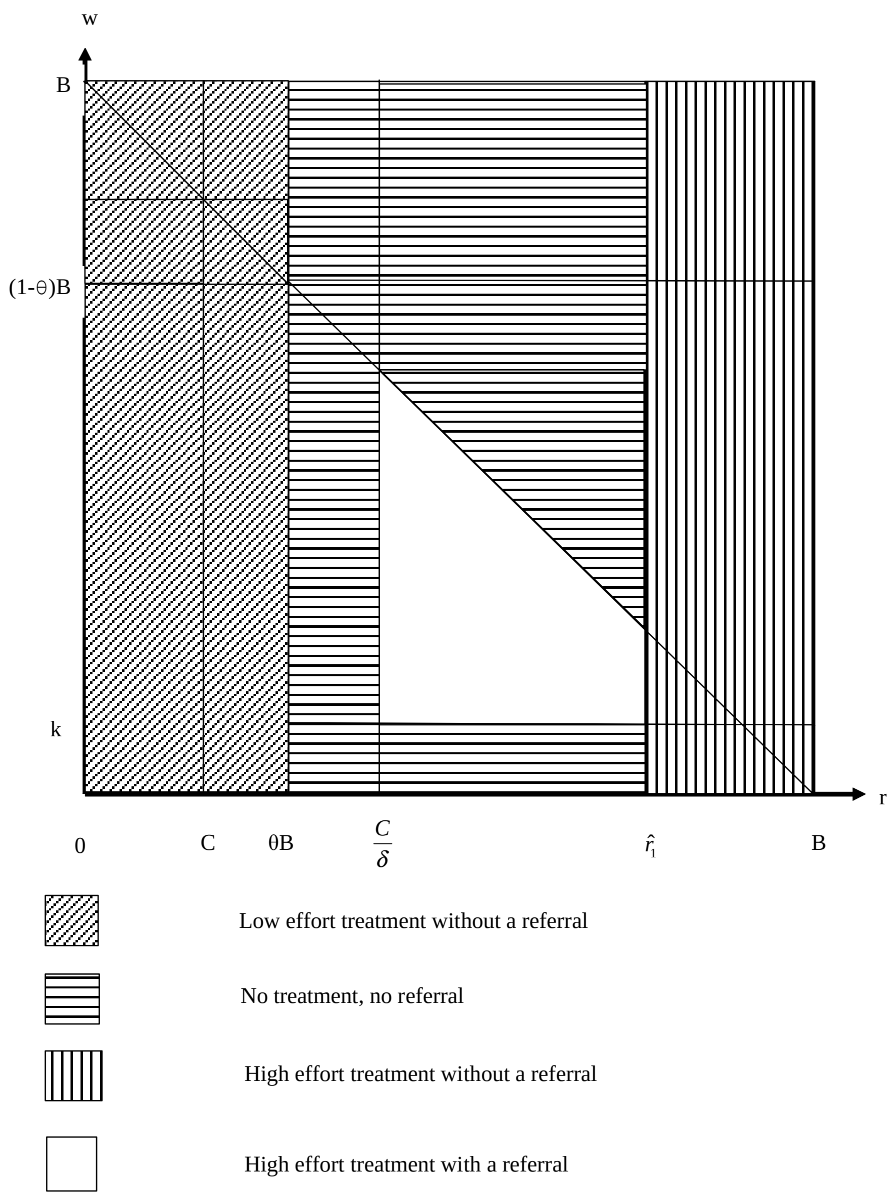


Figure 2.2: Market outcomes when $\frac{C}{\delta} < \theta B < \widehat{r}_1 < B$.

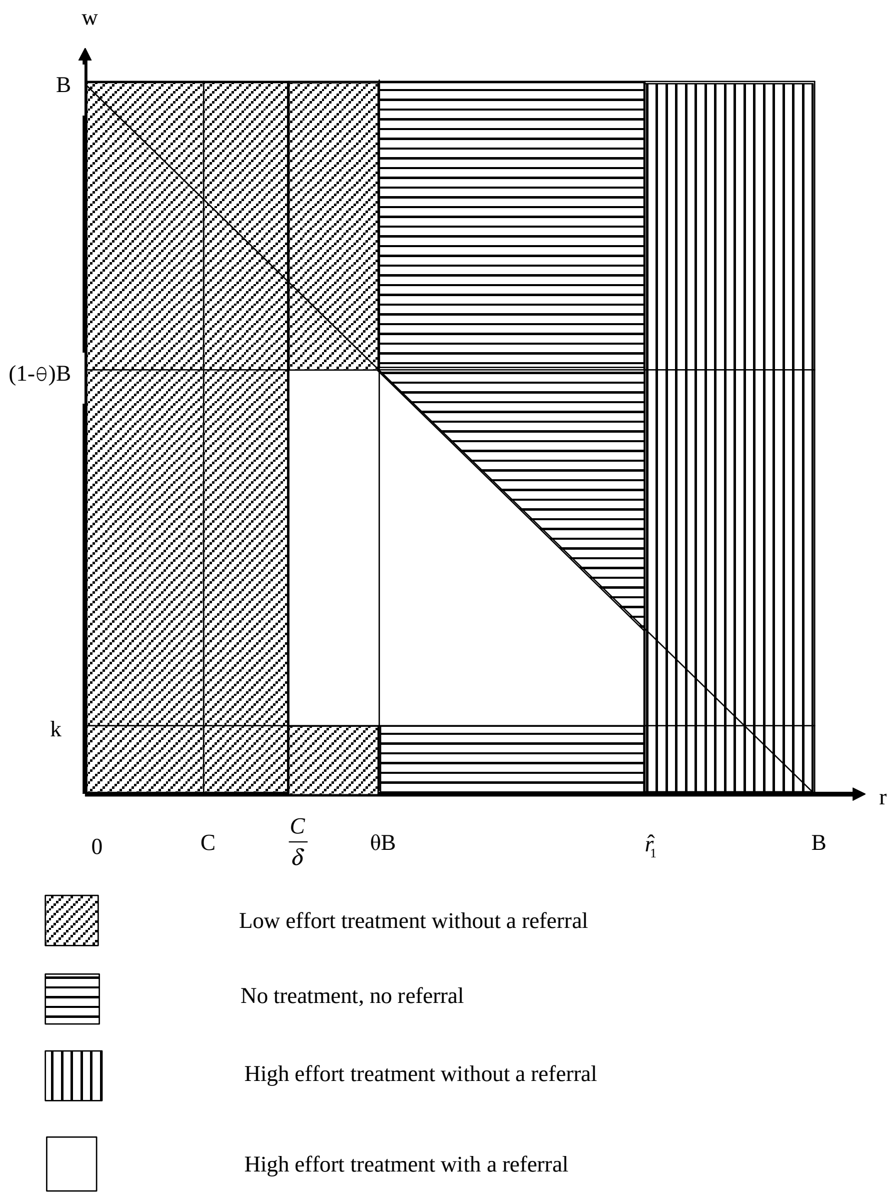
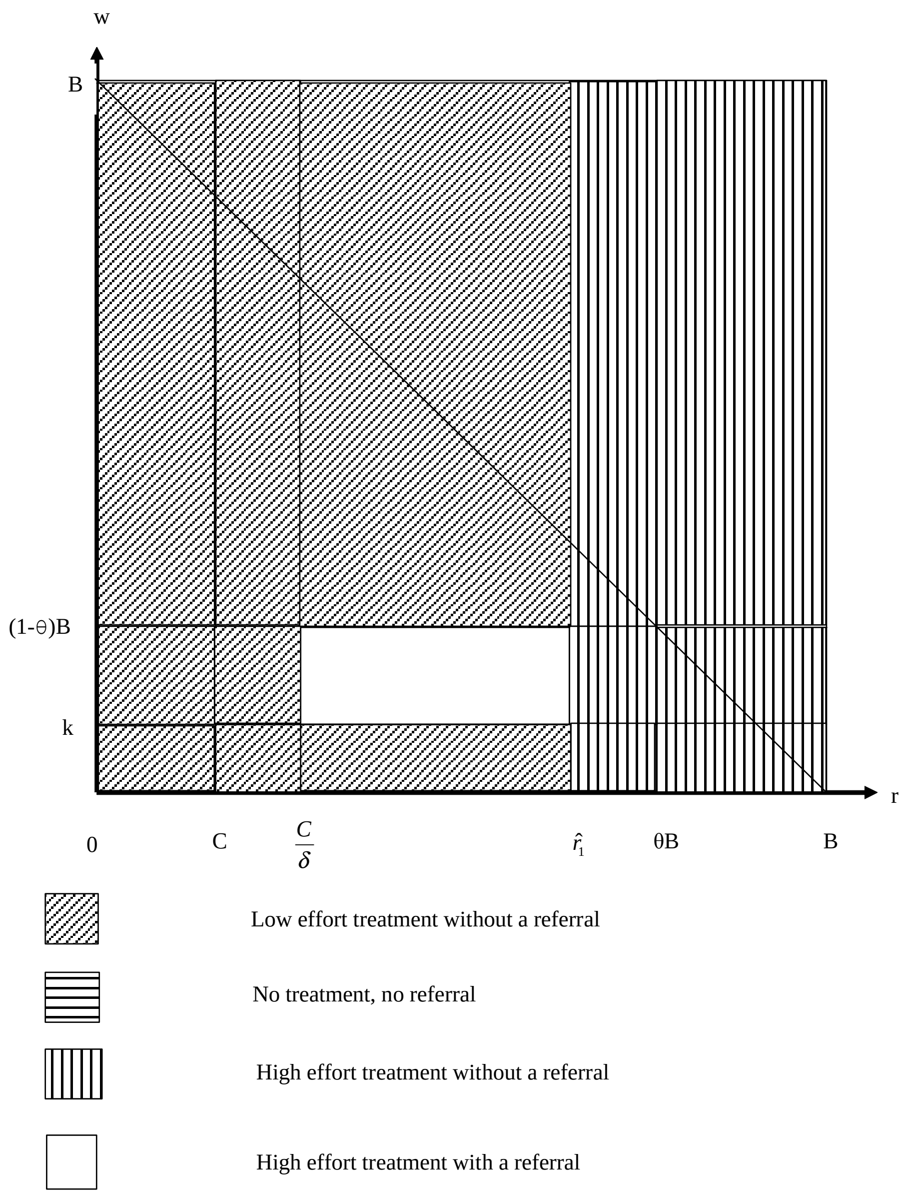


Figure 2.3: Market outcomes when $\frac{C}{\delta} < \widehat{r}_1 < \theta B$.



punish the specialist much more effectively than could any individual patient. The potential loss of future business from a GP with a large patient pool is much more significant than the potential loss of future business from a single patient.

There is a related literature that uses reputation to explain the existence of institutions, primarily firms[15], unions[16] and retailers[17]. This literature builds on earlier work examining the extent to which reputation effects and market forces provide an incentive for parties to exert effort when such effort is costly and unobservable.[18] In this section, we compare the model presented in this paper with other reputation based theories of the existence of institutions. In particular, we compare it with Kreps' theory of firms ([61]), Hogan's theory of unions ([50]) and Biglaiser and Friedman's theory of retailers ([15]).

Kreps ([61]) provides a reputation-based explanation for the existence of firms. His starting point is a static moral hazard problem similar to the one considered in this paper. The outcome of a transaction between a consumer and a producer depends on some action taken by the producer which is unobservable to the consumer at the time of the transaction. For example, a consumer's satisfaction with a product may depend on its quality, which might not be observed until the product is actually used, well after the time of purchase. Furthermore, this quality level may be unverifiable to third parties. If low quality products are cheaper to produce, then the producer may have an incentive to pretend that a low quality product is really a high quality product. However, if the producer has a long-run relationship with the consumer, he runs the risk of losing that consumers future custom if he misleads the consumer about product

[15]See Kreps ([61]) and Tadelis ([121]) in particular.

[16]See Hogan ([50]), MacLeod and Malcolmson ([65]) and Malcolmson ([67]) in particular.

[17]See Biglaiser ([14]) and Biglaiser and Friedman ([15]) in particular.

[18]The important earlier papers include Cooper and Ross ([24]), Diamond ([31]), Holmstrom ([53]), Klein and Leffler ([58]) and Shapiro ([104]). More recent work along these lines can be found in Horner ([54]).

quality. As such, repetition may overcome the static moral hazard problem. Unfortunately, consumers will not always have a long-run relationship with the producer. Kreps shows that if the outcomes of previous transactions can be communicated to future customers, then the fact that any individual customer only has a short-run relationship with the producer is irrelevant. What matters is that the producer can be punished in the future for any current transgressions.

The analysis presented in this paper strengthens the foundations of Kreps model in two ways. Kreps' model assumes that the outcomes of current transactions can be accurately and costlessly communicated to future consumers. It also assumes that the terms of trade between consumers and producers are fixed because of the existence of competition for trading partners. However, Kreps does not explicitly examine either the communication process or the price formation process. In this paper, we have provided a natural means of communicating past outcomes in the form of gatekeeping intermediaries that monitor the outcomes of transactions that result from their referrals. Furthermore, we have explicitly modelled the process of price formation. This has allowed us to provide foundations for the fixed terms of trade assumption and to determine the equilibrium terms of trade.

Hogan ([50]) provides a reputation-based explanation for the existence of unions. He considers a moral hazard problem between a worker and a firm in which output depends on employee effort which is costly for the employee to provide. The effort provided by any given employee is assumed to be observable to both the employee and the employer, but is both unobservable and unverifiable to third parties. As a result, the employer must use an implicit contract to motivate high effort on the part of employees. If the employer and an employee have only a short-run relationship, the employer will have an incentive to renege on any promised high effort payments for employees, even if they provide high effort. This problem can be at least partially overcome if the employer and the employees have a long-run relationship. However, if the firm's production technology exhibits diminishing marginal returns to labour, implicit contracts will be insufficient to achieve first-best employment levels. The presence of a union in this setting increases employment and thereby allows efficiency losses to be reduced. The reason for this is that the union is able to monitor the behaviour of the employer and inform its members if the employer has reneged on a contract with any them. Note that the individual employees cannot undertake this monitoring role themselves because they do not observe the effort choices of other employees. The union is assumed to possess a technology that enables it to observe the effort choices of its members. This technology is too expensive for individual employees to utilise in the absence of the union. The monitoring costs incurred by the union are recovered through union membership fees.

Unlike the model considered in this paper, the employer and the employees have a long-run relationship in Hogan's model. As such, reputation can play a role in reducing the occurrence of opportunistic behaviour by the firm, even if the union is not present. The presence of the union simply enhances the effectiveness of these reputation effects. A gatekeeping intermediary in Hogan's model would look more like a temporary recruitment agency. The very nature of temporary employment would ensure that temporary employees have only a short-run relationship with the employer. As such, in the absence of a temporary recruitment agency, the employer would have a strong incentive to renege on any payments that were promised in return for the provision of high effort by the employee. This would in turn provide an incentive for the employee to only provide low effort. However, if the employer obtains his temporary employees through a temporary recruitment agency, then that agency will have a long-run relationship with the firm. Furthermore, if that agency refers its workers to many different firms, then it will have a long-run relationship with the temporary employees that use its referral services. As such, the temporary recruitment agency may be able to leverage its long-run relationships with employers and temporary employees to create an artificial long-run relationship between the employers and the temporary employees. As such, the use of temporary recruitment agencies may allow for equilibria in which the employees provide high effort and the employers do not renege on their promise to pay extra for the provision of high effort.

Biglaiser and Friedman ([15]) provide a reputation-based theory of the existence of retailers. They consider a moral hazard problem in which consumers do not observe the quality of a good until after they have purchased it. As such, producers will have an incentive to mislead consumers about the quality of the goods they produce. This incentive is reduced if the producer sells his products through retailers that also stock the products of other producers rather than directly to the public. The reason for this is that the retailers will lose future sales on their other products if they do not punish a producer for misleading consumers about the quality of his products.

In many respects, Biglaiser and Friedman's model is the closest in spirit to the one employed in this paper. In both models, an intermediary has a long-run relationship with producers because he sells their products to many different consumers. Similarly, in both models an intermediary has a long-run relationship with consumers because he sells many different products that they might wish to purchase. However, there are also a number of differences. One relatively minor difference relates to the party that chooses to use the services of an intermediary. In Biglaiser and Friedman's model, the producers choose to use an intermediary to market their goods to consumers. In the model employed in this paper, consumers choose to use an intermediary to access the services of a producer.

There are also more significant differences between the two models. Biglaiser and Friedman allow quality to vary continuously, while specialist effort can only take on one of two values in the model considered in this paper. In Biglaiser and Friedman's model, consumers learn the quality of the products they purchase following the transaction. If they purchased the product from an intermediary, the intermediary also learns the quality of the product after the transaction has been completed. As such, there is never any uncertainty about whether or not they have been deceived by the producer. However, in the model employed in this paper, patients and general practitioners cannot always infer the effort choices of specialists. In order to allow for the possibility that producers may mislead consumers about the quality of their products, Biglaiser and Friedman incorporate a signaling component into their model. No signaling components are incorporated into the model considered in this paper. One final difference in the structure of the two models relates to the length of the relationship between consumers and producers. In Biglaiser and Friedman's model, in the absence of retailers, demand in every period will be the same if the producer never defects. The proportional decrease in demand is identical to the proportion of customers that were deceived in the previous period. This is consistent with consumers having a long-run relationship with the producer. The main focus of this paper, on the other hand, is on situations in which patients have only a short-run relationship with specialists.

One of the most significant differences between Biglaiser and Friedman's analysis and the analysis presented in this paper relates to the type of equilibria that are considered. Even when retailers are absent, Biglaiser and Friedman focus on equilibria in which producers do not mislead consumers about the quality of their products. The introduction of retailers reduces the cost to a producer of signalling his chosen quality level and reduces the price that a consumer must pay to receive a product of that quality level. While a similar result was obtained in the model employed in this paper for the case in which both patients and GPs could motivate high effort treatment from specialists, this case was not the main focus of this paper. The main focus of this paper was on situations in which patients could not motivate high effort treatment from specialists, but GPs could do so. We showed that there existed equilibria in which this was the case that were preferred by patients to the outcomes when GPs were not present. Patients preferred this equilibrium because the additional expected benefit they received from high effort treatment exceeded the additional cost of such treatment. In some cases, specialists also preferred the presence of GPs. In these cases, the additional revenue that specialists received from providing high effort treatment exceeded the additional cost of providing high effort treatment.

2.8 Appendix: Industry structure with endogenous prices

Throughout this paper, we have assumed that prices are set exogenously. This assumption can be viewed as a black-box for any price formation process that generates a uniform price. It does not really matter whether health care markets are perfectly competitive, imperfectly competitive or even monopolised, as long as price discrimination is not present. However, in the previous section on industry structure when prices are exogenous, we imposed a slightly stronger assumption. This assumption involved the equilibrium treatment price being the same when GPs are present as it is when they are absent. In this section, we relax this assumption and examine market outcomes in the context of an explicit price formation process. Equilibrium prices are assumed to be the outcome of Bertrand competition. We will allow specialists to offer two different prices, one for high effort treatment and one for low effort treatment. However, since effort is not observable and treatment outcomes are not verifiable, the high effort treatment price will need to satisfy the high effort incentive compatibility constraint if it is to be credible. The difference between the high effort incentive compatibility constraint for patient-specialist interactions and the corresponding constraint for GP-specialist interactions suggests that the equilibrium treatment price may vary with the presence of GPs in some cases under this price formation process.

2.8.1 Price formation without GPs

In the absence of GPs, standard Bertrand competition arguments suggest that the equilibrium high effort treatment price will be

$$r_{11}^{*} = \left[1 + \frac{(1-\delta)}{\pi\delta(1-\theta)}\right] C.$$

This is the lowest price at which patients will believe that specialists will provide high effort treatment. Similarly, standard Bertrand arguments suggest that the equilibrium low effort treatment price will be

$$r_{10}^{*} = 0.$$

Patients will demand high effort treatment if and only if

$$B - \left[1 + \frac{(1-\delta)}{\pi\delta(1-\theta)}\right] C \geq \theta B,$$

which can be rearranged to yield

$$C \leq \left[\frac{\pi\delta\left(1-\theta\right)^2}{\pi\delta\left(1-\theta\right)+\left(1-\delta\right)} \right] B.$$

Thus, if the cost of providing high effort treatment is not too high, then the prevailing treatment price will be r^*_{11}. However, if the cost of high effort is too high, then the prevailing treatment price will be zero.

2.8.2 Price formation with GPs

When GPs are present, we need to determine both the equilibrium treatment price and the equilibrium referral price. In this section, we will allow specialists to offer three types of treatment service. They can offer high effort treatment to patients with a referral, high effort treatment to patients without a referral and low effort treatment. Once again, each of these outcomes needs to be self-enforcing. We have already described the candidate treatment prices in the absence of a referral. As such, we only need consider the case in which a patient seeks both a referral and treatment. If GPs have infinite patient pools, standard Bertrand arguments suggest that the equilibrium treatment price will be

$$r^*_{\infty 1} = \frac{C}{\delta}.$$

Furthermore, assuming there are an infinite number of potential GPs who stand ready to enter at zero cost, Bertrand competition among GPs will result in an equilibrium referral price of $w^* = k$. As such, patients will seek both a referral and high effort treatment if and only if

$$B - \frac{C}{\delta} - k \geq \max\left\{ B - \left[1 + \frac{(1-\delta)}{\pi\delta(1-\theta)} \right] C, \theta B \right\}.$$

If a patient would prefer high effort treatment without a referral to no treatment whatsoever, then this becomes

$$B-\frac{C}{\delta}-k\geq B-\left[1+\frac{(1-\delta)}{\pi\delta(1-\theta)}\right]C,$$

which can be rearranged to obtain

$$k\leq\left\{\frac{(1-\delta)\left[1-\pi\left(1-\theta\right)\right]}{\pi\delta\left(1-\theta\right)}\right\}C.$$

This is equivalent to

$$C\geq\left\{\frac{\pi\delta\left(1-\theta\right)}{(1-\delta)\left[1-\pi\left(1-\theta\right)\right]}\right\}k.$$

On the other hand, if a patient would prefer low effort treatment to high effort treatment without a referral, then the patient will seek high effort treatment with a referral if and only if

$$B-\frac{C}{\delta}-k\geq\theta B,$$

which can be rearranged to obtain

$$k\leq(1-\theta)B-\frac{C}{\delta}.$$

This is equivalent to

$$C\leq\delta\left[\left(1-\theta\right)B-k\right].$$

Thus we know that the prevailing treatment price will be $r^*_{\infty 1}$ if either

$$\left\{\frac{\pi\delta\left(1-\theta\right)}{(1-\delta)\left[1-\pi\left(1-\theta\right)\right]}\right\}k\leq C\leq\left[\frac{\pi\delta\left(1-\theta\right)^2}{\pi\delta\left(1-\theta\right)+(1-\delta)}\right]B,$$

or

$$\left[\frac{\pi\delta\left(1-\theta\right)^{2}}{\pi\delta\left(1-\theta\right)+\left(1-\delta\right)}\right]B \leq C \leq \delta\left[\left(1-\theta\right)B-k\right].$$

2.8.3 Market outcomes with endogenous prices

There are three possible outcomes in this market. The first outcome involves all patients obtaining both a referral and high effort treatment. The second outcome involves all patients obtaining high effort treatment without a referral. The third case involves all patients obtaining low effort treatment without a referral.

All patients will obtain both a referral and high effort treatment if either

$$\left\{\frac{\pi\delta\left(1-\theta\right)}{\left(1-\delta\right)\left[1-\pi\left(1-\theta\right)\right]}\right\}k \leq C \leq \left[\frac{\pi\delta\left(1-\theta\right)^{2}}{\pi\delta\left(1-\theta\right)+\left(1-\delta\right)}\right]B$$

or

$$\left[\frac{\pi\delta\left(1-\theta\right)^{2}}{\pi\delta\left(1-\theta\right)+\left(1-\delta\right)}\right]B \leq C \leq \delta\left[\left(1-\theta\right)B-k\right].$$

In these cases, the equilibrium treatment price will be

$$r_{\infty 1}^{*}=\frac{C}{\delta},$$

while the equilibrium referral price will be

$$w^{*}=k.$$

Suppose that

$$\Omega=\left\{\varpi:\left\{\frac{\pi\delta\left(1-\theta\right)}{\left(1-\delta\right)\left[1-\pi\left(1-\theta\right)\right]}\right\}k \leq \varpi \leq \left[\frac{\pi\delta\left(1-\theta\right)^{2}}{\pi\delta\left(1-\theta\right)+\left(1-\delta\right)}\right]B\right\}$$

and

$$\Psi = \left\{ \psi : \left[\frac{\pi\delta\left(1-\theta\right)^2}{\pi\delta\left(1-\theta\right)+\left(1-\delta\right)} \right] B \leq \psi \leq \delta\left[\left(1-\theta\right)B-k\right] \right\}.$$

Note that it is possible that Ω might be an empty set. Similarly, it is possible that Ψ might be an empty set. In order for patients not to obtain a referral, we need both $C \notin \Omega$ and $C \notin \Psi$.

All patients will obtain high effort treatment without a referral if both $C \notin \Omega \cup \Psi$ and

$$C \leq \left[\frac{\pi\delta\left(1-\theta\right)^2}{\pi\delta\left(1-\theta\right)+\left(1-\delta\right)} \right] B.$$

In this case, the equilibrium treatment price will be

$$r_{11}^* = \left[1 + \frac{(1-\delta)}{\pi\delta(1-\theta)} \right] C,$$

while the referral market will not exist.

All patients will obtain low effort treatment without a referral if both $C \notin \Omega \cup \Psi$ and

$$C > \left[\frac{\pi\delta\left(1-\theta\right)^2}{\pi\delta\left(1-\theta\right)+\left(1-\delta\right)} \right] B.$$

In this case, the equilibrium treatment price will be

$$r_{10}^* = 0,$$

while the referral market will not exist.

Chapter 3

A learning theory of referrals

3.1 Introduction

The potential for adverse selection problems to result in market failure is well understood.[1] Nonetheless, there are a number of ways in which the parties to a transaction might be able to reduce the impact of adverse selection. The informed party in a transaction might attempt to signal his type. Alternatively, the uninformed party might attempt to design a screening contract that induces the informed party to truthfully reveal his type. However, in a static setting, such mechanisms might only be partially successful.[2] If the transacting parties interacted repeatedly, then reputation effects might potentially reduce the

[1] See, for example, Akerlof ([3]), Pauly ([82]), Rothschild and Stiglitz ([96]), Stiglitz and Weiss ([119]) and Wilson ([127]).

[2] Useful surveys of the signalling and screening literature include Hirshleifer and Riley ([49], Chapter 11), Kreps ([60], Chapter 17), Mas-Colell et al ([68], Chapter 13), Riley ([92]) and Stiglitz and Weiss ([120]). The seminal papers in this literature include Banks and Sobel ([11]), Cho and Kreps ([20]), Riley ([90], [91]), Rothschild and Stiglitz ([96]), Spence ([114]) and Wilson ([127]).

incidence of adverse selection below the level that would occur in a static setting. Unfortunately, there are many occasions in which parties to a transaction do not repeatedly interact. In the absence of repeated interaction, reputation cannot be relied upon to overcome adverse selection. As such, we might expect adverse selection problems to be particularly severe in markets characterised by few or infrequent interactions between the trading parties. Is consumer protection regulation the only safeguard available in such settings or can institutions be devised that might capture the benefits of repeated interaction?

A gated industry structure might provide one possible solution to potential adverse selection problems between parties that interact infrequently. An industry has a gated structure when consumers seek a referral to a producer from an intermediary, rather than accessing the services of a producer directly. These intermediaries aggregate many short-run transactions between various consumers and a particular producer. This might enable an intermediary to learn the producer's level of proficiency more rapidly than an individual consumer. A gated industry structure is observed in some professional service industries, including the medical and legal professions in some countries. In the medical industries of some Commonwealth countries, including Australia, it is unusual for patients to visit a specialist without first obtaining a referral from a general practitioner (GP).[3] Indeed, while it is permissible for a patient to be treated by a specialist without a referral in Australia, there are financial incentives offered to patients who obtain a referral for treatment by a specialist. In most medical specialties, patients will be reimbursed a larger portion of their treatment costs under the Medicare system in Australia if they obtain a referral before seeking treatment.[4] This raises two interesting questions. Why do we observe a gated structure for

[3] Commonwealth countries in which many patients obtain a referral from a general practitioner before seeking the services of a specialist include Australia ([83], p. 421; [23], part 2, p. 3), New Zealand ([97], section 2, p. 14) and the United Kingdom ([19]).

[4] See, for example, Commonwealth of Australia ([23], part 2, p. 3).

some professional service industries? Why might government regulation be required to support this gated industry structure? In this paper we provide an answer to both of these questions. As an aid to exposition, we will focus on the health care example.

This paper is organised as follows. First, we outline a competitive model of a health care market in which treatment outcomes depend on the ability of the treating specialist, which is private information. We analyse the outcomes in a static version of this health care market. In particular, we characterise the conditions under which the market will fail to exist because of the adverse selection problem. We then proceed to show that if this static market is repeated an infinite number of times, the resulting dynamic market is less likely to fail to exist than the static market. Following this, we consider the impact of introducing a gated structure to the dynamic health care market. This further reduces the potential for the health care market to fail to exist. A comparison between these three versions of the health care market is then provided. This comparison illustrates the benefits of both repetition and the presence of intermediaries. Following this, we show that the benefits of the gated structure might not be achievable without government intervention. The reason for this is the presence of a positive information externality. Finally, we conclude by comparing the results of this chapter with those that we obtained in Chapter 2 and with some features of actual health care markets.

3.2 A competitive model of health care markets

Consider an economy with three groups of agents who live forever. These groups are patients, general practitioners (GPs) and medical specialists. Let patients be indexed by $i \in \{1, 2, \cdots, I\}$, GPs by $j \in \{1, 2, \cdots, J\}$ and specialists by $k \in \{1, 2, \cdots, K\}$. We will assume that there are an infinite number of patients

($I \to \infty$) and specialists ($K \to \infty$), but only a finite numbers of GPs ($J < \infty$). Patients are either well ($d = 0$) or sick ($d = 1$) In each period, a patient is randomly a disease state, $d \in \{0, 1\}$. Following this, each patient can choose whether or not to seek treatment if he is sick. Treatment can sometimes result in a cure, improving the patient's health status for that period. The probability that a sick patient is cured by treatment increases with the ability of the treating specialist. Patients can seek a referral to the specialist from a GP if they believe that this will increase the probability that they are treated by a high ability specialist. Both referrals and treatments come at a price. For budget constrained patients, the benefits of an increased probability of good health need to be weighed against the foregone consumption of other goods that expenditure on health care entails. We will assume that patients visit neither a GP nor a specialist when they are healthy.[5]

All agents in this economy are price takers who behave as though the existing prices are exogenously specified. We will focus on stationary equilibria for this economy, so that prices don't change over time. The price per referral from any GP is w, while the price per treatment from a medical specialist is r. We will assume throughout that specialist ability is neither observable nor verifiable by outside parties, although it may be learned by patients and GPs that interact with the specialist. As such, the treatment price cannot vary with specialist ability.

For payoff purposes, time is assumed to be discrete in this economy. Time periods are indexed by $t \in \{0, 1, 2, \cdots\}$, with payoffs occurring at the end of each period. In each period, the market opens and the agents interact within the market. Note that not all agents move at once in the market. The market process involves sequential moves by various agents. Thus the timing of the

[5] If the equilibrium prices for referrals and treatment are positive, this assumption is not needed. Even if these prices are zero, we could avoid making this assumption by introducing an opportunity cost of time (perhaps in the form of foregone leisure) into the model.

moves in the market process is important. We will maintain the assumption that time is discrete and index time within a period by $s \in \{0, 1, 2, \cdots\}$. In this fashion, each point in time can be given a unique time stamp of the form $(t, s) \in \{0, 1, 2, \cdots\}^2 = \mathbb{Z}_+^2$.

3.2.1 The timing of the market in each period

Prior to the beginning of the first stage game, at time $t = -1$, Nature selects an ability level for each potential specialist as a sequence of independent draws from a common distribution. Any given specialist has either high ability ($\lambda = 1$) or low ability ($\lambda = 0$). Each specialist's ability level, which is fixed for all time, is observed by nobody except for that specialist. However, it is common knowledge that the probability that any given specialist has a high level of ability is given by $\mu \in (0, 1)$.

At the beginning of each period ($s = 0$), Nature randomly chooses a disease state for each patient, $d_i \in \{0, 1\}$. The disease state for each patient is chosen as a random draw from some common distribution, Π. The probability that any given patient is sick in any given period is $\pi \in (0, 1)$, while the probability that a patient is well in any given period is $(1 - \pi)$. The distribution from which these disease states are drawn is common knowledge.

At $s = 1$, having observed their disease state, patients choose whether or not to seek treatment and, if they seek treatment, whether or not to seek a referral from their GP. If they seek a referral, they choose which GP to visit at $s = 2$. At $s = 3$, GP's choose the specialists to which they will refer their patients. At this point in the stage game, any patients who chose top self-refer at $s = 2$ will also choose the specialist from whom they will treatment. We will assume that GPs follow up on the outcomes from treatment of any of the patients they refer. In this fashion, the GP knows the entire history of outcomes for each of his previous referrals at the start of each period. This allows him to use this information when making his current referral decisions.

Following this, at $s = 4$, specialists treat each patient that has been referred to them. Finally, at $s = 5$, Nature chooses whether or not each patient is cured. If a patient is cured, he will have good health in that period ($h = 1$), while if the patient is not cured, he will have bad health ($h = 0$). We will assume that treatment by a high ability specialist always results in a patient being cured, while treatment by a low ability specialist never results in a patient being cured. Furthermore, any patient who chose not to seek treatment will not be cured.

3.2.2 Player objectives

Every agent in this game is assumed to maximise the discounted present value of a sequence of per-period von Neuman Morgernstern expected utility functions. Furthermore, they all have a common rate of time preference, represented by the stationary discount factor $\delta \in [0, 1)$. Thus differences in the preferences of the three groups of agents arise from differences in their per-period preferences. These are outlined below.

Patients

Patients all have identical per-period preferences defined over their expenditure on health care (p) and their health state (h). These preferences may be represented by a quasi-linear per-period Bernoulli utility function of the form

$$u(h_t, p_t) = B(h_t) - p_t,$$

where $B(0)$ is normalised to zero and $B(1) = B > 0$.

The health state in each period is a random variable and may vary across patients. It depends on whether or not the patient is sick, whether or not treatment is sought and, if so, the ability of the treating specialist. If a sick patient receives treatment from a high ability specialist, he will definitely be cured. If a sick patient receives treatment from a low ability specialist, he will definitely not be cured. Since each patient knows his disease status before having to make any decisions about treatment, the probability of good health in period t is given by

$$\theta_t = \begin{cases} 1 & \begin{array}{c}\text{if either } d = 0 \text{ or} \\ \text{high ability treatment is received with certainty when } d = 1;\end{array} \\ \mu & \begin{array}{c}\text{if } d = 1 \text{ and treatment is sought} \\ \text{from a specialist whose ability is not known;}\end{array} \\ 0 & \begin{array}{c}\text{if } d = 1 \text{ and either low ability treatment is received} \\ \text{or no treatment is sought.}\end{array} \end{cases}$$

Expenditure on health care in any given period may also vary across patients. It will depend on whether or not the patient seeks treatment and, if so, whether or not the patient also seeks a referral. We will assume that patients do not seek a referral if they do not also desire treatment. Thus a patient's expenditure on health care is given by

$$p_t = \begin{cases} 0 & \text{if neither treatment nor referral is sought;} \\ r & \text{if treatment is sought without a referral;} \\ w + r & \text{if both treatment and referral are sought.} \end{cases}$$

This allows us to express a patient's per-period expected utility as

$$Eu(h_t, p_t) = \theta_t B - p_t.$$

General practitioners

GPs are assumed to be risk-neutral. As such, they maximise their expected profits. The Bernoulli utility function that represents their per-period preferences is simply their per-period profit. Let $n_{j,1,t}$ denote the number of referrals a particular GP makes in period t. Assuming that they have a constant marginal cost of C_{GP} per referral and no fixed costs, their per-period profits are

$$\Pi_j = n_{j,1}(w - C_{GP}),$$

where we have dropped the time subscripts for convenience.

Note that $n_{j,1,t}$ is a random variable if a GP has a finite patient pool of size $n_{j,t}$. However, when GP has an infinite patient pool, this source of uncertainty disappears. The reason for this is that $n_{j,1,t}$ converges almost surely to $\pi n_{j,t} = \infty$ as $n_{j,t} \to \infty$.

Proposition 19 *$n_{j,1,t}$ converges almost surely to $\pi n_{j,t} = \infty$ as $n_{j,t} \to \infty$.*

Proof. First, note that

$$n_{j,1} = \frac{n_{j,1}}{n_j} n_j = \alpha_j n_j.$$

Furthermore,

$$\alpha_j = \frac{n_{j,1}}{n_j} = \frac{1}{n_j} \sum_{i(j)=1}^{n(j)} 1_{i(j),1},$$

where $1_{i(j),1}$ is an indicator variable that takes the value of one if the patient in question is sick and the value zero otherwise. Note that GP j has a patient pool consisting of n_j patients, including those that do not need the GPs services in the current period. These patients are indexed by $i_j \in \{1, 2, \cdots, n_j\}$. Note that each of these indicator variables is a Bernoulli random variable that takes on the value one with probability π and zero otherwise. As such, $\left\{1_{i(j),1}\right\}_{i(j)=1}^{n(j)}$ is a sequence of independent and identically distributed Bernoulli random variables in which

$$E\left(1_{i(j),1}\right) = \pi(1) + (1-\pi)(0) = \pi \text{ for all } i_j \in \{1, 2, \cdots, n_j\}.$$

Thus, from the strong law of large numbers[6], we know that

$$\Pr\left(\lim_{n(j)\to\infty} \alpha_j = \pi\right) = 1.$$

This allows us to conclude that α_j converges almost surely to π. Finally, note that

$$\lim_{n(j)\to\infty} n_j = \infty.$$

Thus we can conclude that

$$n_{j,1} = \alpha_j n_j \xrightarrow{a.s.} (\pi)(\infty) = \infty.$$

Hence we know that $n_{j,1,t}$ converges almost surely to $\pi n_{j,t} = \infty$ as $n_{j,t} \to \infty$.

■

Medical Specialists

Like GPs, medical specialists are assumed to be risk-neutral. As such, a specialist's per-period Bernoulli utility function is simply his profit. This profit will depend on the treatment price (r) and the cost of constant marginal cost

[6] See Billingsley ([16], pp. 85-86) for a discussion of the strong law of large numbers. Note that when GP patient pools are finite, the number of members in a GP's patient pool is an integer. As such, when we take the limit as this number approaches infinity, we are restricting our attention to the set of natural numbers. In effect, there is a one-to-one correspondence between each member of a GP's patient pool and each element of the set of natural numbers when that GP has an infinite patient pool. As such, each GP has a countable number of patients. This ensures that the standard version of the strong law of large numbers applies in the model considered in this paper. If we had assumed that each GP had a continuum of patients instead of a countably infinite number of patients, we would have needed to use the techniques mentioned in Judd ([57]). The reason for this is that each GP would have had an uncountable number of patients in that case.

of treatment (C_S). We will assume that there are no fixed costs of treatment. Each specialist's per-period per-patient profit is given by $r - C_S$. Let $n_{k,t}$ denote the number of patients a particular specialist treats in period t. Given this, the specialist's profit in period t is simply

$$\Pi_{k,t} = n_{k,t}(r - C_S).$$

In each period, specialists observes the number of patients seeking treatment from them before actually treating any patients. As such, the only uncertainty that affects medical specialists relates to the number of patients that will seek their treatment services in future periods. While this may be a function of the outcomes that result from their current and past treatment of patients, there is nothing they can do to influence these outcomes. Thus specialists will simply maximise their per-period profits.

3.3 Static outcomes in competitive health care markets

Before analysing the dynamic model of competitive health care markets, it is useful to consider what would happen in the absence of any repetition whatsoever. To do this, we will initially assume that patients can be afflicted with the disease at most once. As such, patients will only need the services of medical specialist at most once. We will focus on a representative patient (i) whose has the disease and a representative specialist (k). We will assume that all agents in this economy are price takers and that all prices are exogenously determined.

Specialist will receive the treatment price (r) from each patient that they treat. However, they will also incur a treatment cost equal to C_S for each patient that they treat. This treatment cost is independent of their ability.

Proposition 20 *(The participation constraint for specialists): Specialists will offer their treatment services if and only if* $r \geq C$.

Proof. Medical specialists are profit maximisers. They can guarantee themselves zero profits by refusing to treat any patients. As such, they will only treat patients if the profit per patient is at least zero. This requires that the treatment price either matches or exceeds the cost of treatment for each patient. ■

Now consider a sick patient. If the patient is cured, then he will be in good health ($h = 1$) for the remainder of the current period. This will yield him benefits equal to $B(h) = B(1) = B > 0$. If the patient is not cured, he will be in bad health ($h = 0$) for the remainder of the current period. This yields him benefits equal to $B(h) = B(0) = 0$. If the patient is to have any chance of being cured of this disease, he will require treatment by a medical specialist. There are two types of medical specialists, high ability specialists ($\lambda = 1$) and low-ability specialists ($\lambda = 0$). If the patient is treated by a high ability specialist, he is guaranteed to be cured. If the patient is treated by a low ability specialist, he is guaranteed not to be cured. Unfortunately, each specialist's ability is private information, known only to that specialist. It is common knowledge, however, that the probability of any given specialist having high ability is $\mu \in (0, 1)$.

Proposition 21 *(The participation constraint for patients): Patients will seek treatment if and only if* $r \leq \mu B$.

Proof. Since the participation constraint for specialists is independent of their ability, patients with rationally believe that the probability of a cure following treatment is equal to the probability that a specialist has high ability. As such, if a patient obtains treatment, his expected utility is:

$$EU_i(treatment) = \mu B + (1 - \mu)0 - r = \mu B - r.$$

If a patient does not obtain treatment, he will neither be cured nor will he have to pay the treatment price. As such, his expected utility will be zero. Thus a patient will seek treatment if and only if $\mu B - r \geq 0$. This requires that $r \leq \mu B$. ■

The outcomes in this static health care market will vary with the treatment price. If the price is too low, no specialists will offer their treatment services. As such, no patients will be cured. If the treatment price is too high, no patients will seek treatment and hence no patients will be cured. In both cases, patients receive zero expected utility and specialists receive zero profit. There will sometimes be an intermediate range of prices in which all patients will seek treatment and all specialists will offer their treatment services. In these cases, patients will receive non-negative expected utility and specialists will earn non-negative profits. However, some patients will be disappointed with the outcome of their treatment. These are the patients that will have been unfortunate enough to be treated by a low ability specialist.

Proposition 22 *(Static market existence): If $\mu B < C_S$, then the set of prices at which both patients demand treatment and specialists supply treatment is empty. If $\mu B \geq C_S$, then the set of prices at which both patients demand treatment and specialists supply treatment is non-empty.*

Proof. Both the specialist participation constraint and the patient participation constraint are satisfied if and only if $C_S \leq r \leq \mu B$. This clearly requires that $C_S \leq \mu B$. Thus, if $C_S > \mu B$, there are no values for the treatment price that will satisfy both participation constraints. If $C_S = \mu B$, then there is unique value for the treatment price that will satisfy both participation constraints. This value is $r = C_S = \mu B$. Finally, if $C_S < \mu B$, then there is a range of values for the treatment price that will satisfy both participation constraints. These are $r \in [C, \mu B]$. ■

Proposition 23 *(Static market outcomes): If the health care market exists, then patients will receive non-negative expected utility and specialists will make non-negative profits. However, some patients might not be cured following treatment.*

Proof. If the health care market exists, then $C_S \leq r \leq \mu B$. This means that

$$EU_i = \mu B - r \geq \mu B - \mu B = 0.$$

It also means that

$$\Pi_{k(d)} = r - C_S \geq C_S - C_S = 0.$$

Thus patients receive non-negative expected utility and specialists receive non-negative profits. However, since both high ability specialists and low ability specialists are prepared to offer their treatment services for this range of prices, some patients might have sought treatment from a low-ability specialist. Any such patients will not be cured. ■

3.4 Dynamic outcomes without general practitioners

Suppose we now play an infinitely repeated version of the stage game in the absence of GPs. In this version of the dynamic model, any patient that wants to be treated by a specialist needs to seek the services of a specialist without a referral. Before observing his disease state in period zero, a patient's lifetime expected utility is

$$J_i = \sum_{t=0}^{\infty} \delta^t \pi EU_{i,t}(sick) + \sum_{t=0}^{\infty} \delta^t (1-\pi) EU_{i,t}(well).$$

The expected utility that patients receive in periods when they are well is not affected by their treatment choices when they are sick. As such, we can ignore these terms when considering the impact of treatment decisions on a patient's lifetime expected utility. Given this, for the remainder of this paper we will only focus on the payoffs that a patient receives in periods when he is sick.

The first time a patient is afflicted with the disease, he will not have any information about the ability of any of the specialists. As such, he may as well randomly select a specialist from whom to seek treatment if he decides to seek treatment.

Proposition 24 *(Treatment payoff): If it is optimal for a patient to seek treatment when he is first afflicted with the disease, then it will be optimal for him to seek treatment whenever he is afflicted with the disease. Furthermore, his lifetime expected utility, net of at that point in time is*

$$V = \left[\frac{(1-\delta)+\delta\pi}{(1-\delta)^2+(1-\delta)\mu\delta}\right]\mu B - \left[\frac{(1-\delta)+\mu\delta\pi}{(1-\delta)^2+(1-\delta)\mu\delta}\right]r.$$

Proof. The patient does not know the ability of whichever specialist treats him the first time he is afflicted with the disease. As such, his expected utility in that period is simply $\mu B - r$. Following treatment, the patient is either cured or not cured. As such, he learns the ability of the treating specialist. If he is cured, then he knows the treating specialist has a high level of ability. If it is ever optimal to seek treatment, then it must be optimal to do so when you know that you will be cured. As such, the patient will seek treatment from that specialist whenever he is sick in future periods. If the patient is not cured in the current period, then he knows the treating specialist has a low level of ability. As such, he will never seek treatment from that specialist again. However, he might still get sick in some future periods. In the next such period, he will need to start from scratch. Since there an infinite number of potential medical specialists, the problem he will face in that period will be identical to the current one. As such, the lifetime expected utility from that period onwards will be identical to the lifetime expected utility in the current period. Hence, the patient's lifetime expected utility the first time he gets sick, prior to treatment, must satisfy the following equation:

$$V = (\mu B - r) + \mu \sum_{t=1}^{\infty} \delta^t \pi (B - r) + (1 - \mu) \sum_{t=1}^{\infty} \delta^t \pi (1 - \pi)^{t-1} V.$$

Note that this equation can be rewritten as

$$V = (\mu B - r) + \mu\delta \sum_{t=0}^{\infty} \delta^t \pi (B - r) + (1 - \mu)\delta \sum_{t=0}^{\infty} \left[\delta \left(1 - \pi\right)\right]^t \pi V,$$

which becomes

$$V = (\mu B - r) + \mu\delta \left(\frac{\pi(B - r)}{1 - \delta} \right) + (1 - \mu)\delta \left(\frac{\pi V}{1 - \delta + \delta\pi} \right).$$

Solving this equation for the patient's lifetime expected utility (V) yields:

$$V = \left[\frac{\left(1 - \delta + \delta\pi\right)^2}{\left(1 - \delta + \delta\pi\mu\right)\left(1 - \delta\right)} \right] \mu B - \left[\frac{\left(1 - \delta + \delta\pi\right)}{\left(1 - \delta\right)} \right] r.$$

■

While we have characterised a patients lifetime expected utility if he chooses to seek treatment, we have not yet established the conditions under which seeking treatment will be optimal. In order to derive these conditions, we first need to lifetime expected utility of a sick patient who does not seek treatment.

Proposition 25 *(Non-treatment payoff): If it is optimal for a patient not to seek treatment when he is first afflicted with a the disease, then it will never be optimal for him to seek treatment. Furthermore, his lifetime expected utility at that point in time is zero.*

Proof. If the patient does not seek treatment when he is first afflicted with the disease, then he will neither be cured in that period nor incur any medical expenses in that period. As such, his expected utility in that period is zero. Furthermore, since he did not seek treatment, he will not learn anything about the ability of any of the medical specialists. As such, he will face an identical problem the next time he gets sick, assuming such an event occurs. If it is optimal for him not to seek treatment the first time he gets sick, it must therefore be optimal for him not to seek treatment the next time he gets sick. Thus, by mathematical induction, if it is not optimal for a patient to seek treatment the first time he is sick, it will never be optimal for him to seek treatment when he is sick. The lifetime expected utility of such a patient will be zero. ■

Now that we have characterised the payoffs to a patient, both when he chooses to seek treatment and when he does not, we are in a position to derive the conditions under which it will be optimal for him to seek treatment when he is sick.

Proposition 26 *(Participation constraint for patients): A sick patient will seek treatment in this dynamic health care market if and only if*

$$r \leq \left[\frac{(1-\delta+\delta\pi)}{(1-\delta+\delta\pi\mu)}\right]\mu B.$$

Proof. We have already established that a patient will either always seek treatment when he sick or never do so. Furthermore, we have calculated the lifetime expected utility at the point in time where the patient first discovers that he sick for both of these cases. Thus we know that a sick patient will choose to seek treatment if and only if his lifetime expected utility from doing so is at least as large as his lifetime expected utility from not seeking treatment. This requires that

$$\left[\frac{(1-\delta+\delta\pi)^2}{(1-\delta+\delta\pi\mu)(1-\delta)}\right]\mu B-\left[\frac{(1-\delta+\delta\pi)}{(1-\delta)}\right]r\geq 0.$$

This inequality can be rearranged to obtain

$$r\leq\left[\frac{(1-\delta+\delta\pi)}{(1-\delta+\delta\pi\mu)}\right]\mu B.$$

Thus patients will seek treatment whenever the treatment price is not too high.

■

It is worth noting that the range of treatment prices for which patients will be willing to seek treatment in this dynamic health care market is larger than that in a static health care market. This makes intuitive sense because the benefits from seeking treatment in the dynamic market are larger than they are in the static market. As well as the expected benefits from treatment in the current period, which occur in both markets, patients who seek treatment will also learn the ability level of the treating specialist. While this has no value in a static model, it yields positive expected utility in a dynamic market.

Proposition 27 *(The benefits of repetition): The maximum treatment price at which a patient will seek treatment is higher in a dynamic market than it is in a static market.*

Proof. Let $\widehat{r}_d$ denote the maximum price at which patients will seek treatment in a dynamic market and $\widehat{r}_s$ denote the maximum price at which patients will seek treatment in a static market. These prices are given by least upper bounds of the patient participation constraints in each model. Note that

$$\widehat{r}_d-\widehat{r}_s=\left[\frac{(1-\delta+\delta\pi)}{(1-\delta+\delta\pi\mu)}\right]\mu B-\mu B,$$

which can be rearranged to obtain

$$\widehat{r}_d - \widehat{r}_s = \left[\frac{(1-\mu)\delta\pi}{(1-\delta+\delta\pi\mu)}\right]\mu B > 0.$$

Thus we know that $\widehat{r}_d > \widehat{r}_s$. ■

There is nothing that specialists can do to influence treatment outcomes, treatment costs or treatment prices. This means that repetition does not affect their participation decisions. As such, the participation constraint facing specialists in dynamic health care markets will be the same as that facing them in static health care markets. We are now in a position to characterise the conditions under which a dynamic health care market will exist. As with a static health care market, the outcomes in this dynamic health care market will vary with the treatment price. If the price is too low, no specialists will offer their treatment services. As such, no patients will be cured. If the treatment price is too high, no patients will seek health care treatment and hence no patients will be cured. In both cases, patients receive zero expected utility and specialists receive zero profit. There will sometimes be an intermediate range of prices in which all patients will seek treatment and all specialists will offer their treatment services. In these cases, patients will receive non-negative lifetime expected utility and specialists will earn non-negative profits. However, some patients will be disappointed with the outcome of their treatment on at least one occasion. These are the patients that will have been unfortunate enough to be treated by a low ability specialist in at least one period.

Proposition 28 *(Dynamic market existence): If $\widehat{r}_d < C_S$, then the set of prices at which both patients demand treatment and specialists supply treatment is empty. If $\widehat{r}_d \geq C_S$, then the set of prices at which both patients demand treatment and specialists supply treatment is non-empty.*

Proof. Both the specialist participation constraint and the patient participation constraint are satisfied if and only if $C_S \leq r \leq \widehat{r}_d$. This clearly requires that $C_S \leq \widehat{r}_d$. Thus, if $C_S > \widehat{r}_d$, there are no values for the treatment price that will satisfy both participation constraints. If $C_S = \widehat{r}_d$, then there is unique value for the treatment price that will satisfy both participation constraints. This value is $r = C_S = \widehat{r}_d$. Finally, if $C_S < \widehat{r}_d$, then there is a range of values for the treatment price that will satisfy both participation constraints. These are $r \in [C_S, \widehat{r}_d]$. ■

Proposition 29 *(Dynamic market outcomes): If the health care market exists, then patients will receive non-negative expected utility and specialists will make non-negative profits. However, some patients might not be cured following treatment.*

Proof. If the health care market exists, then the participation constraints of both patients and specialists must be satisfied. Thus patients receive non-negative expected utility and specialists receive non-negative profits. However, since both high ability specialists and low ability specialists are prepared to offer their treatment services for this range of prices, some patients might have sought treatment from a low-ability specialist on at least one occasion when they were sick. Any such patients will not have been cured on those occasions.
■

We showed earlier that the repetition present in this dynamic health care market expands the set of treatments prices for which patients will be willing to seek treatment compared to the set of such prices in a static health care market. Since the set of treatment prices for which specialists will offer their services is the same in both markets, this means that there is a larger set of circumstances in which a dynamic health care market will exist than in which a static health care market will exist.

Proposition 30 *(The relationship between static market existence and dynamic market existence): A dynamic health care market will exist whenever a static health care exists. Furthermore, a static health care market will not exist whenever a dynamic health care market does not exist. However, there are some cases where a dynamic health care market will exist but a static health care market will not exist.*

Proof. First, we will show that the existence of a static health care market implies the existence of a dynamic health care market. We have already established that a static health care market will exist if and only if $C_S \leq \widehat{r}_s$. We have also already established that $\widehat{r}_d > \widehat{r}_s$. Thus, if a static health care market exists, we know that $C_S < \widehat{r}_d$. Thus the the condition that guarantees the existence of a dynamic health care market ($C_S \leq \widehat{r}_d$) is satisfied. Thus the existence of a static health care market does indeed imply the existence of a dynamic health care market. Now we show that the non-existence of a dynamic health care market implies the not existence of a static health care market. If a dynamic health care market does not exist, then we know that $C_S > \widehat{r}_d$. Since $\widehat{r}_d > \widehat{r}_s$, this means that $C_S > \widehat{r}_s$ as well. This means that, if a dynamic health care market cannot exist, then nor can a static health care market. Finally, we will show that there are some cases in which a dynamic health care market will exist, but a static health care market will not exist. Suppose that $\widehat{r}_s < C_S \leq \widehat{r}_d$. In this case, the necessary and sufficient condition for the existence of a static health care market is not satisfied, but the necessary and sufficient condition for the existence of a dynamic health care market is satisfied. Hence there are cases where a dynamic health care market can exist but a static health care market cannot exist. ■

3.5 Dynamic outcomes with general practitioners

Having established what happens in a dynamic health care market without general practitioners, we are now in a position to analyse the impact of introducing them. Suppose, for the moment, that access to treatment by medical specialists is subject to regulation. Specifically, we will assume that all sick patients will need to seek a referral before obtaining treatment in the first period (period zero). In all subsequent periods, sick patients will be able to choose whether or not to seek a referral before obtaining treatment. A regulation along these lines is needed because of the presence of an information externality. This issue is discussed in more detail later in this paper. Market outcomes in period zero will be deferred until later in this paper as well. In this section, we will focus on outcomes in this form of a dynamic health care market after period zero has finished.

Suppose that there are a finite number of GPs, each of whom has an infinite patient pool in period zero of this dynamic health care market. In these circumstances, each GP must have an infinite number of sick patients in period zero.

Proposition 31 *(Patient numbers): Each GP has and in infinite number of sick patients in period zero.*

Proof. We showed earlier that $n_{j,d}$ converges almost surely to $\pi_d n_j = \infty$ as $n_{j,t} \to \infty$. Given that all GPs have infinite patient pools in period zero, they will almost surely have an infinite number of sick patients in that period. ■

Since each GP has an infinite number of sick patients in period zero, he can refer a single patient to each of an infinite number of medical specialists. As such, every GP will find at least one medical specialist who has high ability.

Proposition 32 *(GP learning): In period zero, every GP will find at least one medical specialist who has high ability.*

Proof. Recall that any given specialist has high ability with probability μ and low ability with probability $(1-\mu)$. Furthermore, recall that GPs observe the treatment outcomes for all of the patients for whom they provide a referral. Since the ability level of each specialist is perfectly revealed by the outcome of any treatment that they provide, a GP will learn the ability level of any specialist to whom he refers at least one patient. Since the GP has an infinite number of sick patients, he can refer a single patient to each of an infinite number of medical specialists. The probability that at least one of these specialists has high ability is simply one minus the probability that none of the specialists who treat a patient referred by the GP has high ability. This is given by

$$\Pr\left\{\#_{j,k}(\lambda=1)>0\right\} = 1 - \lim_{n(j,1)\to\infty}\left(\mu^{n(j,1)}\right) = 1-0=1.$$

Thus, in period zero, each GP will find at least one medical specialist who has high ability. ■

Thus it is possible for every GP to find at least one high ability specialist in period zero. Since the ability of specialists is fixed for all time prior to the opening of the dynamic health care market in period zero, every GP can guarantee a patient that he will be cured if he seeks a referral from that GP in any time period after period zero.

A potential problem with referrals is that GPs might have an incentive to refer patients to a low ability specialist. The reason for this is that patients will no longer need a referral after they learn the identity of a high ability specialist. If GPs are earning positive profits on each referral they make, they might attempt to induce further demand for services by initially making poor referrals. However, patients can deter such a strategy by threatening to dump any GP who refers them to a low ability specialist after period zero. This will remove any incentive that GPs might have to make poor referrals after period zero.

Proposition 33 *(GP incentive compatibility constraint): General practitioners cannot profit by referring patients to specialists who have low ability after period zero.*

Proof. If all consumers employ a strategy that involves never again using the referral services of a GP who refers them to a low ability specialist after period zero, then GPs will not gain anything by referring any patient to a low ability specialist after period zero.. As such, after period zero, GPs will be indifferent between referring patients to high ability specialists and referring them to low ability specialists. In these circumstances, it is reasonable to assume that GPs will refer patients to high ability specialists after period zero. ■

In order to analyse market existence and market outcomes under these circumstances, we need to consider the behaviour of two groups of consumers. The first group of consumers are those that are fortunate enough to have been treated by a high ability specialist in period zero. The second group of consumers consists of all consumers who are not members of group one. This includes both patients who were not sick in period zero, patients who were sick in period zero but were unfortunate enough to be treated by a low ability specialist and patients who were sick but chose not to seek treatment in that period. The lifetime utility after period zero will be different for these two groups of patients. The first time a patient in group one gets sick after period zero, his lifetime expected utility if he seeks treatment will be

$$V_{i,1} = (B - r) + \sum_{t=1}^{\infty} \delta^t \pi (B - r) = \left(\frac{1 - \delta + \delta\pi}{1 - \delta} \right) (B - r) .$$

If a patient in group two seeks a referral after period one, he knows that he will be referred to a high ability specialist. As such, he will learn the identity of a high ability specialist. Thus, the first time a patient in group two gets sick after period zero, his lifetime expected utility if he seeks a referral and treatment is

$$V_{i,2} = (B - r - w) + \sum_{t=1}^{\infty} \delta^t \pi (B - r) = \left(\frac{1 - \delta + \delta\pi}{1 - \delta} \right) (B - r) - w.$$

With these lifetime expected utilities in hand, we can characterise the circumstances under which members of each will choose to seek treatment when they are sick.

Proposition 34 *(Participation constraint for informed patients): A patient who knows the identity of a high ability medical specialist will treatment whenever he is sick if and only if $r \leq B$. Furthermore, if $r > B$ he will never seek treatment.*

Proof. The remaining lifetime expected utility of a sick patient who knows the identity of a high ability specialist is given by $V_{i,1}$. Furthermore, this is true for any period in which he is sick. This patient will seek treatment when he is sick if and only if $V_{i,1} \geq 0$. This requires that $r \leq B$. Thus a patient who knows the identity of a high ability specialist will seek treatment whenever he is sick if and only if $r \leq B$. ■

Proposition 35 *(Participation constraint for uniformed patients who have access to an informed GP): A patient who does not know the identity of a high ability specialist but whose GP does know the identity of such a specialist will weakly prefer to seek both a referral and treatment over no treatment when he is first sick if and only if*

$$r \leq B - \left(\frac{1 - \delta}{1 - \delta + \delta\pi} \right) w.$$

Furthermore, if such a patient seeks both a referral and treatment the first time he is sick and $w \geq 0$, he will seek treatment whenever he is sick from that point in time onwards.

Proof. The remaining lifetime expected utility of a sick patient who does not knows the identity of a high ability specialist but whose GP does know the identity of a high ability specialist is given by $V_{i,1}$ if that patient seeks both a referral and treatment. This patient will prefer to seek both a referral and treatment if and only if $V_{i,1} \geq 0$. This requires that

$$\left(\frac{1-\delta+\delta\pi}{1-\delta}\right)(B-r)-w \geq 0,$$

which can be rearranged to yield

$$r \leq B-\left(\frac{1-\delta}{1-\delta+\delta\pi}\right)w.$$

Finally, note that following the referral and treatment, such a patient will know the identity of a high ability specialist. We know that informed patients will be willing to seek treatment whenever they are sick if $r \leq B$. Thus if $w \geq 0$ and uninformed patients with access to an informed GP seeks a referral after period zero, then it must the case that $r \leq B$. ■

Recall that repetition does not affect the participation decision of specialists. As such, the participation constraint facing specialists in this dynamic health care markets with GPs will be the same as that facing them in a static health care market. We are now in a position to characterise the conditions under which a dynamic health care market with GPs will exist.

Proposition 36 *(Treatment market existence): If all sick patients seek treatment in period zero, then the treatment market will exist after period zero when-*

ever $C_S \leq B$. However, some patients will choose not to seek treatment if

$$r \in \left(\max \left\{ C_S, \left[\frac{(1-\delta+\delta\pi)}{(1-\delta+\delta\pi\mu)} \right] \mu B, B - \left(\frac{1-\delta}{1-\delta+\delta\pi} \right) w \right\}, B \right].$$

Proof. Recall that patients are not allowed to seek treatment without a referral in period zero. As such, if every sick patient in period zero seeks treatment, then every sick patient will also seek a referral. Since each GP has an infinite patient pool, this means that each GP will almost surely have an infinite number of sick patients in period zero. We have already shown that this ensures that each GP will be able to discover the identity of at least one high ability specialist. Since a GP only discovers the identity of a high ability specialist when at least one of his patients is treated by a high ability specialist, we know that at least one patient from the patient pool of each GP must also learn the identity of a high ability specialist. This means that there will be at least J informed patients at the end of period zero. If these patients are ever sick from the beginning of period one onwards, they will seek treatment whenever $r \leq B$. Specialists will be willing to provide a referral if and only if $r \geq C_S$. As such, the treatment market will exist if and only if $C_S \leq B$. On the other hand, if $C_S > B$, then the treatment market will not exist. The mere existence of the treatment market does not mean that all sick patients will choose to seek treatment. Clearly informed patients will seek treatment. However, uninformed patients might not do so. Recall that uninformed patients would be willing to seek treatment without a referral if and only if

$$r \leq \left[\frac{(1-\delta+\delta\pi)}{(1-\delta+\delta\pi\mu)} \right] \mu B = \left[\frac{\mu(1-\delta)+\delta\pi\mu}{(1-\delta)+\delta\pi\mu} \right] B < B.$$

Furthermore, assuming that $w \geq 0$, uninformed patients will be willing to seek treatment with a referral if and only if

$$r \leq B - \left(\frac{1-\delta}{1-\delta+\delta\pi} \right) w < B.$$

Thus uninformed patients will only participate in the treatment market if

$$\max \left\{ \left[\frac{(1-\delta+\delta\pi)}{(1-\delta+\delta\pi\mu)} \right] \mu B, B - \left(\frac{1-\delta}{1-\delta+\delta\pi} \right) w \right\} \geq C_S.$$

If the treatment market exists but this condition does not hold, then it must be the case that

$$\max \left\{ C_S, \left[\frac{(1-\delta+\delta\pi)}{(1-\delta+\delta\pi\mu)} \right] \mu B, B - \left(\frac{1-\delta}{1-\delta+\delta\pi} \right) w \right\} = C_S$$

and $r \in [C_S, B]$. This means that uninformed patients will not participate in the treatment market despite the fact that it exists. ■

Since GPs are also present in the health care sector now, we also need to consider the existence of a referral market. This requires us to examine participation constraints for patients and GPs in the referral market after period zero.

Proposition 37 *(Patient participation constraint for referrals): Informed patients will never seek a referral. Uninformed patients will seek a referral if and only if both*

$$r \leq B - \left(\frac{1-\delta}{1-\delta+\delta\pi} \right) w$$

and

$$w \leq \left(\frac{1-\delta+\delta\pi}{1-\delta+\delta\pi\mu} \right) (1-\mu) B.$$

Proof. Recall that

$$\widehat{r}_d = \left[\frac{(1-\delta+\delta\pi)}{(1-\delta+\delta\pi\mu)} \right] \mu B.$$

Let

$$\widehat{r}_{GP} = B - \left(\frac{1-\delta}{1-\delta+\delta\pi} \right) w.$$

Consider an uniformed patient who is sick at some time after period zero. If this patient seeks both a referral and treatment, his remaining lifetime expected utility will be

$$V_i(referral) = \begin{cases} \left(\frac{1-\delta+\delta\pi}{1-\delta}\right)(B-r) - w & \text{if} \quad r \leq \widehat{r}_{GP}; \\ \\ 0 & \text{if} \quad r > \widehat{r}_{GP}. \end{cases}$$

If the patient seeks treatment without a referral, his remaining lifetime expected utility will be

$$V_i(treatment) = \begin{cases} \left[\frac{(1-\delta+\delta\pi)^2}{(1-\delta+\delta\pi\mu)(1-\delta)}\right]\mu B - \left[\frac{(1-\delta+\delta\pi)}{(1-\delta)}\right] r & \text{if} \quad r \leq \widehat{r}_d; \\ \\ 0 & \text{if} \quad r > \widehat{r}_d. \end{cases}$$

Finally, if the patient seeks neither treatment nor referral, his remaining lifetime expected utility will be zero. A patient will prefer to seek to seek a referral and treatment to no treatment whatsoever if and only if

$$\left(\frac{1-\delta+\delta\pi}{1-\delta} \right) (B-r) - w \geq 0.$$

This can be rearranged to obtain

$$r \leq B - \left(\frac{1-\delta}{1-\delta+\delta\pi} \right) w.$$

Furthermore, a patient will prefer to seek both a referral and treatment over treatment alone if and only if $V_i(referral) \geq V_i(treatment)$. If $r \leq \min\{\widehat{r}_d, \widehat{r}_{GP}\}$, this requires that

$$\left(\frac{1-\delta+\delta\pi}{1-\delta}\right)(B-r)-w\geq\left[\frac{(1-\delta+\delta\pi)^2}{(1-\delta+\delta\pi\mu)(1-\delta)}\right]\mu B-\left[\frac{(1-\delta+\delta\pi)}{(1-\delta)}\right]r.$$

This can be rearranged to obtain

$$w\leq\left(\frac{1-\delta+\delta\pi}{1-\delta+\delta\pi\mu}\right)(1-\mu)B.$$

■

Proposition 38 *(GP participation constraint): GPs will offer their referral services whenever* $w \geq C_{GP}$.

Proof. We have already established that patients can deter GPs from referring them to low ability specialists. As such, repetition does not affect the participation decision of GPs. Recall that there are no fixed referral costs and constant marginal referral costs. As such, the average cost of a referral is constant. Indeed, it is simply the marginal referral cost. Hence GPs will participate if and only if the referral price exceeds this referral cost. This requires that $w \geq C_{GP}$.

■

With the participation constraints for patients and GPs in hand, we are now in a position to establish the conditions under which a referral market will exist.

Proposition 39 *(Referral market existence): A referral market will exist if and only if*

$$C_{GP}\leq\left(\frac{1-\delta+\delta\pi}{1-\delta+\delta\pi\mu}\right)(1-\mu)B.$$

Proof. Note that only uninformed patients will seek a referral after period zero. These patients will seek a referral if and only if

$$w\leq\left(\frac{1-\delta+\delta\pi}{1-\delta+\delta\pi\mu}\right)(1-\mu)B.$$

GPs will offer their referral services if and only if $w \geq C_{GP}$. As such, the set of referral prices for which both patients and GPs are willing to participate in the referral market will be non-empty if and only if

$$C_{GP} \leq \left(\frac{1-\delta+\delta\pi}{1-\delta+\delta\pi\mu} \right) (1-\mu) B.$$

■

Finally, we are in a position to comment on the impact that the introduction of GPs has on the threshold treatment price for market existence.

Proposition 40 *(Treatment price when a referral market exists): If the referral market exists, then the threshold treatment price will satisfy the following condition:*

$$\widehat{r}_{GP} \in \left[\widehat{r}_d, B - \left(\frac{1-\delta}{1-\delta+\delta\pi} \right) C_{GP} \right] \subseteq [\widehat{r}_d, B].$$

Proof. Recall that, for the referral market to exist, we need

$$C_S \leq w \leq \left(\frac{1-\delta+\delta\pi}{1-\delta+\delta\pi\mu} \right) (1-\mu) B.$$

As such, we know that

$$\widehat{r}_{GP} \leq B - \left(\frac{1-\delta}{1-\delta+\delta\pi} \right) C_{GP} \leq B.$$

Furthermore, we also know that

$$\widehat{r}_{GP} \geq B - \left(\frac{1-\delta}{1-\delta+\delta\pi} \right) \left(\frac{1-\delta+\delta\pi}{1-\delta+\delta\pi\mu} \right) (1-\mu) B,$$

which can be rearranged to obtain

$$\widehat{r}_{GP} \geq \left[\frac{(1-\delta+\delta\pi)}{(1-\delta+\delta\pi\mu)} \right] \mu B = \widehat{r}_d.$$

■

3.6 A comparison of treatment market outcomes

We can now compare the various circumstances in which a treatment market will exist and the treatment outcomes in each of these circumstances. Recall that a static treatment market will exist if $C_S \leq r \leq \mu B = \widehat{r}_s$, while a dynamic treatment without GPs will exist if

$$C_S \leq r \leq \left[\frac{(1-\delta+\delta\pi)}{(1-\delta+\delta\pi\mu)}\right]\mu B = \widehat{r}_d.$$

Furthermore, a dynamic treatment market with GPs in which all sick patients seek treatment will exist if

$$C_S \leq r \leq B - \left(\frac{1-\delta}{1-\delta+\delta\pi}\right)C_{GP} = \widehat{r}_{GP}.$$

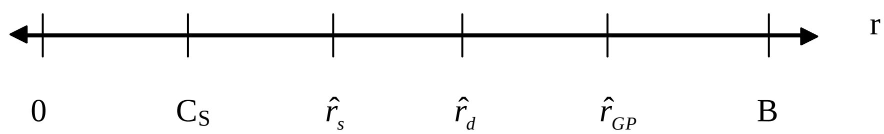
Finally, a dynamic treatment market with GPs in which only informed patients seek treatment will exist if

$$B - \left(\frac{1-\delta}{1-\delta+\delta\pi}\right)C_{GP} = \widehat{r}_{GP} \leq r \leq B.$$

The relationship between these various threshold treatment prices is illustrated in Figure 3.1. It is clear that repetition expands the range of treatment prices for which the treatment market will exist. Repetition and the presence of GPs expands the range of treatment prices that are consistent with treatment market existence even further.

Treatment market outcomes will vary with the nature of the market and the prevailing prices. In both the static market and the dynamic market without GPs, all patients will seek and obtain treatment whenever the treatment price

Figure 3.1: The relationship between threshold treatment prices



lies between the cost of treatment and the relevant threshold treatment price. However, in both cases, patients will not know whether or not they will be cured. Some patients will be cured, while some patients will not be cured. In a dynamic market with GPs, all patients will seek treatment if the treatment price lies between the cost of treatment and the threshold price for uninformed patients. However, if the treatment price lies between the threshold treatment price for uninformed patients and the benefit from good health, then only the informed patients will seek treatment. In both of these cases, all of the patients who seek treatment will be cured.

3.7 Information externalities and the need for regulation

While characterising the outcomes in dynamic markets with GPs, we assumed that all sick patients were required to seek a referral and treatment in period zero. This allowed every GP to learn the identity of at least one high ability specialist before the start of the next period. In this section, we will examine patients referral choices in period zero. We will show that all sick patients choosing to seek a referral in that period is not an equilibrium outcome. The intuition for this result involves the presence of a positive information externality. Individual patients bear the entire cost of obtaining a referral. However, they do not capture any of the benefits from that referral. The patient would learn the ability level of one of the specialists if he sought treatment regardless of whether or not he also sought a referral. The benefit from the referral is that the GP also learns the ability level of one of the specialists for each referral that he makes. Given the presence of this positive externality, it is not surprising that patients may choose to consume too few referrals in period zero, from a social welfare point of view. As such, some policy to correct for this may be warranted. Potential policies include a requirement that patients seek a referral before obtaining treatment in period zero or some form of subsidy for patients who seek a referral. As we noted in the introduction to this chapter, the cost of treatment is often subsidised for patients who seek a referral in Australia.

Proposition 41 *(The need for regulation): All sick patients in period zero choosing to seek a referral is not part of an equilibrium outcome if referrals are not free.*

Proof. Suppose that all but one of the sick patients in period zero choose to seek both a referral and treatment. Consider the choice confronting the remaining sick patient. There are an infinite number of sick patients in period zero who choose to seek a referral from each GP regardless of this patient's referral decision. As such, every GP will still be able to learn the identity of at least one high ability GP at the end of period zero. Furthermore, the probability of this patient becoming an informed patient at the end of period zero is not altered by his referral choice. As such, the continuation payoff for this patient will not be altered by his referral decision. This means that he can simply maximise his period zero expected utility when making a referral decision. The expected utility for this patient in period zero if he seeks both a referral and treatment is

$$EU_i(referral) = \mu B - r - w.$$

The expected utility for this patient in period zero if he seeks only treatment is

$$EU_i(treatment) = \mu B - r.$$

Clearly

$$EU_i(referral) \geq EU_i(treatment),$$

with the inequality being strict if $w > 0$. As such, this patient will choose not to seek a referral in period zero if referrals are not free. Thus it cannot be an equilibrium outcome for all sick patients to seek both a referral and treatment in period zero if referrals are not free. ■

It appears that sick patients will not seek referrals in sufficient numbers to allow GPs to become informed at the end of period zero. As such, some form of regulation will be needed if each GP is to be able to learn the identity of at least one high ability specialist. Suppose that any sick patient who wants to obtain treatment in period zero is required to obtain a referral prior to seeking treatment. Will such patients still choose to obtain treatment in period zero?

Proposition 42 *(Patient participation constraint for period zero): If sick patients are required to obtain a referral before seeking treatment in period zero, then they will seek treatment if and only if*

$$w \leq \left(\frac{1-\delta+\delta\pi}{1-\delta+2\delta\pi}\right)\left[\left(\frac{\mu-\delta\mu+\delta\pi}{1-\delta}\right)B-\left(\frac{1-\delta+\delta\pi}{1-\delta}\right)r\right].$$

Proof. The lifetime expected utility of a sick patient in period zero who chooses to seek both a referral and treatment can be decomposed into three terms. These terms are the patient's expected utility in period zero, his continuation utility if he becomes informed at the end of period zero and his continuation utility if he does not become informed at the end of period zero. The continuation payoffs will also need to be weighted by the probability of their occurrence. We will assume throughout that all of the other sick patients during period zero choose to seek a referral. Since we are deriving a condition under which this will be true, this assumption will be valid if that condition holds. The patient's expected utility in period zero is

$$EU_i(referral) = \mu B - r - w.$$

His continuation utility if learns the identity of a high ability specialist during period zero is

$$V_{i,1} = \sum_{t=1}^{\infty} \delta^t \pi(B - r) = \frac{\delta\pi(B - r)}{(1 - \delta)}.$$

His continuation utility if he does not learn the identity of a high ability specialist during period zero is

$$V_{i,2} = \sum_{t=1}^{\infty} \delta^t \pi(B - r) - \sum_{t=1}^{\infty} \delta^t (1 - \pi)^{t-1} \pi w.$$

This expression simplifies to

$$V_{i,2} = \frac{\delta\pi(B - r)}{(1 - \delta)} - \frac{\delta\pi w}{(1 - \delta + \delta\pi)}.$$

Thus the patients lifetime expected utility if he seeks a referral is

$$V_i = \mu B - r - w + \mu \left(\frac{\delta\pi(B - r)}{(1 - \delta)} \right) + (1 - \mu) \left(\frac{\delta\pi(B - r)}{(1 - \delta)} - \frac{\delta\pi w}{(1 - \delta + \delta\pi)} \right),$$

which can be rearranged to obtain

$$V_i = \left(\frac{\mu - \delta\mu + \delta\pi}{1-\delta}\right) B - \left(\frac{1-\delta+\delta\pi}{1-\delta}\right) r - \left(\frac{1-\delta+2\delta\pi}{1-\delta+\delta\pi}\right) w.$$

Clearly, the patient will choose to seek a referral if and only if $V_i \geq 0$, which requires that

$$\left(\frac{\mu - \delta\mu + \delta\pi}{1-\delta}\right) B - \left(\frac{1-\delta+\delta\pi}{1-\delta}\right) r - \left(\frac{1-\delta+2\delta\pi}{1-\delta+\delta\pi}\right) w \geq 0.$$

This expression can be rearranged to obtain

$$w \leq \left(\frac{1-\delta+\delta\pi}{1-\delta+2\delta\pi}\right) \left[\left(\frac{\mu - \delta\mu + \delta\pi}{1-\delta}\right) B - \left(\frac{1-\delta+\delta\pi}{1-\delta}\right) r\right].$$

■

Finally, we need to establish the conditions under which the treatment and referral markets will exist in period zero.

Proposition 43 *(Treatment market existence in period zero): If there is a regulation requiring any patient that wants treatment in period zero to obtain a referral as well, then the treatment market will only exist if*

$$C_S \leq \left(\frac{\mu(1-\delta)+\delta\pi}{1-\delta+\delta\pi}\right) B - \left(\frac{(1-\delta)(1-\delta+2\delta\pi)}{(1-\delta+\delta\pi)^2}\right) w.$$

Proof. Recall that specialists will offer their treatment services if and only if $r \geq C_S$. We can rearrange the period zero participation constraint for patients to obtain

$$r \leq \left(\frac{\mu(1-\delta)+\delta\pi}{1-\delta+\delta\pi}\right) B - \left(\frac{(1-\delta)(1-\delta+2\delta\pi)}{(1-\delta+\delta\pi)^2}\right) w.$$

Thus the participation constraints for patients and GPs can be simultaneously satisfied if and only if

$$C_S \leq \left(\frac{\mu(1-\delta) + \delta\pi}{1-\delta+\delta\pi} \right) B - \left(\frac{(1-\delta)(1-\delta+2\delta\pi)}{(1-\delta+\delta\pi)^2} \right) w.$$

■

Proposition 44 *(Referral market existence in period zero): Even with a regulation requiring any patient that wants treatment in period zero to obtain a referral as well, the referral market will only exist if*

$$C_{GP} \leq \left(\frac{1-\delta+\delta\pi}{1-\delta+2\delta\pi} \right) \left[\left(\frac{\mu - \delta\mu + \delta\pi}{1-\delta} \right) B - \left(\frac{1-\delta+\delta\pi}{1-\delta} \right) r \right].$$

Proof. If this condition is not satisfied, then the participation constraints for patients and GPs cannot be simultaneously satisfied. ■

3.8 Conclusion

Some professional service industries display a gated structure, notably including the medical industry in some Commonwealth countries. The main focus of this chapter has been on explaining both the existence of gatekeeping intermediaries who refer consumers to one of many ultimate producers and providing a rationale for regulations that encourage the use of referrals. The results in this chapter are complementary to those obtained in Chapter 2. In that chapter, the gated industry structure observed in some professional service industries provided an artificial long-run relationship between patients and specialists when, in the absence of GPs, they would only have a short-run relationship. The artificial long-run relationship between patients and specialists enabled them to avoid a market failure resulting from shirking on the part of specialists. This industry structure was largely driven by the demands of patients, although there may have been some circumstances in which the presence of GPs improved the welfare of both patients and specialists. While we provided an explanation for the gated structure of some professional service industries in Chapter 2, that explanation did not provide a rationale for regulations that encouraged such a structure. In this chapter, we have provided a rationale for such regulations. However, while patients in the model employed in Chapter 2 had an incentive to repeatedly seek a referral for the treatment of non-chronic diseases, patients in the model employed in this chapter will seek a referral at most twice. The reason for this difference relates to the underlying market failure. In Chapter 2, the underlying market failure is a moral hazard problem. Specialists can alter their effort choices from period to period. As such, they constantly need to be induced to provide high effort treatment. In this chapter, the underlying market failure is an adverse selection problem. The ability level of a specialist is private information, known only by that specialist. However, this ability level is fixed for all time. Thus, if a patient learns the identity of a high effort specialist, he will obtain no additional benefits from seeking further referrals.

In actual health care markets, patients might well seek a referral on a number of occasions. As such, it would appear that the limited number of referrals that are predicted by the model in this chapter is somewhat unrealistic. However, that result is generated by the stationary population of agents in the model employed in this paper. In actual health care markets, the populations of patients, GPs and specialists will be in a constant state of flux. In each period, some new agents will arrive and some old agents will leave. Thus we would expect the outcomes in actual health care markets to reflect aspects of both the period zero outcomes and the later period outcomes of the model employed in this chapter.

Chapter 4

Multiple interactions and the management of local commons

4.1 Introduction

It is often claimed that private property rights are essential for the efficient performance of an economy. Secure private property rights allow economies to avoid resource misallocations due to externalities and provide the investment incentives necessary to promote economic growth.[1] Despite these claims, there is evidence that some externalities can be managed without the use of private property rights. In particular, various examples of the successful community management of local common property resources have been documented.[2] This

[1] See, for example, Alchian and Allen ([4], pp. 91-96 and 345-347), Coase ([21]), De Long and Shleifer ([30]), Gordon ([41]), Grafton et al ([42]), Hardin ([46]), Johnson et al ([56]), Murphy et al ([74]) and Scully ([101]).

[2] See, for example, Bardhan ([12]), Cornes and Sandler ([25], pp. 283-289), Dietz et al ([32], 2003), Ostrom et al ([79]), Ostrom and Gardner ([80]), Pretty ([86]) and Seabright ([102]).

chapter provides a model in which small communities that have limited interaction with the outside world are able to manage local commons in circumstances where communities that are either larger or have more interaction with the outside world could not manage them. One implication of these results is that, rather than being a necessary precursor to development, the need for private property rights might be generated by the process of development.

This chapter proceeds as follows. First, we set up a simple static model of a village economy. This model takes the form of a static common property resource game and a set of identical static games that represent the other interactions among villagers. Following this, we analyse an infinitely repeated version of the simplified village economy. This allows us to examine how the ability of the village to successfully manage the local commons varies with the number of non-anonymous interactions among the villagers. We then describe the nature of economic development for this village economy and discuss the impact of economic development on the ability of the village to manage its local commons without the use of private property rights or some other form of explicit regulation. Finally, we conclude by discussing some potential extensions of the model presented in this chapter.

4.2 A simple model of a village economy

Imagine a small village in which there are I residents. Suppose that the village has no contact with the outside world, so that it is a closed economy. We are focussing on such a village because we want there to be a relatively large number of non-anonymous transactions between the villagers. The residents of this village want to consume $(n + 1)$ final goods and services. This includes n excludible public goods[3] and one private good. The production of the private

[3] A useful survey of the literature on public goods is contained in Cornes and Sandler ([25]). The seminal references on public goods include Samuelson ([99], [100]). In this paper,

good requires both labour services and the use of a common property resource. The production of the public goods requires only labour services. The villagers combine to produce the public goods. The public goods in this model capture the non-anonymous transactions that would occur in a small, isolated village. The private good provides a potential incentive for over-exploitation of the common property resource. A formal model of a village economy with these features is described in the appendix to this paper.

In the main body of the paper, we will work with a simplified version of this village economy model. Our focus will be on symmetric pure-strategy Nash equilibria. As such, we can represent the incentives facing an individual villager in this economy through the use of two payoff matrices. One of these payoff matrices relates to the interaction of the villagers through their use of the common property resource to produce the private good. The other payoff matrix summarises the interaction among villagers in the production of each of the public goods.

If all other villagers choose a common pure strategy, then the payoffs for villager i in the common property resource interaction may be represented by the payoff matrix in Table 4.1, where $M > W > Z > X > 0$:

Table 4.1: Payoffs in the common property resource interaction

	Cooperate (C_{-i})	Don't Cooperate(D_{-i})
Cooperate (C_i)	W	X
Don't Cooperate (D_i)	M	Z

There are n other interactions between the villagers in each period. Each of

the excludible public goods are simply proxies for the non-anonymous interactions between the various members of a small community.

these interactions can be represented by the payoff matrix in Table 4.2, where $2V > C > V > 0$:

Table 4.2: Payoffs in other interactions

	Cooperate (C_{-i})	Don't Cooperate(D_{-i})
Cooperate (C_i)	$V - \frac{C}{I}$	$V - C$
Don't Cooperate (D_i)	0	0

These payoff matrices summarise the payoffs to an individual villager if all of the other villagers employ a common strategy. As such, they are only a subset of all of the potential payoffs in this village economy game. Since the underlying game is symmetric, however, these two payoff matrices provide all of the relevant payoffs for the calculation of symmetric pure-strategy Nash equilibria in the underlying game. Nonetheless, it is useful to briefly consider the nature of the underlying game. The underlying game consists of $(n + 1)$ simultaneous-move games. Furthermore, each of these games is conducted simultaneously. One of these games represents the interaction of the villagers through their use of the common property resource. This game is an I-player version of the classic Prisoner's dilemma game. This game very neatly captures the incentive structure underlying a static version of a common-property resource problem. The other n games represent the interaction of the villagers in the production of the public goods. There is one game for each public good and each of these games is identical. Recall that the villagers can exclude individual villagers from consuming the services of a public good in this model. This removes the usual incentive for each villager to shirk on his contribution of labour services to the provision of public goods. We will assume that the cost of providing the public good exceeds the benefit that any individual villager obtains from its services. As such, the public good interaction can be modelled as a coordination game in which there are two pure-strategy Nash equilibria. One of these equilibria involves cooperation among all villagers in the production of the public good and hence the provision of the public good. The other equilibria involves the non-cooperation of all villagers in the production of the public good and hence the non-provision of the public good. The cooperative equilibrium Pareto-dominates the noncooperative equilibrium in this public goods game.

4.2.1 A two person example of the village economy game

The common property resource interaction in each period is represented by the stage game in Table 4.3, where $Y > W > Z > X > 0$ and $2W > X + Y$:

Table 4.3: Two player common property resource interaction

	Cooperate (C_{-i})	Don't Cooperate(D_{-i})
Cooperate (C_i)	W, W	X, Y
Don't Cooperate (D_i)	Y, X	Z, Z

There are n other interactions between the villagers in each period. Each of these interactions can be represented by the stage game in Table 4.4, where $2V > C > V > 0$:

Table 4.4: Two player other interactions

	Cooperate (C_{-i})	Don't Cooperate(D_{-i})
Cooperate (C_i)	$V - \frac{C}{2}, V - \frac{C}{2}$	$V - C, 0$
Don't Cooperate (D_i)	$0, V - C$	$0, 0$

There are two pure strategy Nash equilibria in the two-person version of this village economy game. One of these equilibria involves the villagers cooperating in the provision of the public goods and not cooperating in the exploitation of the common property resource. We will call this the cooperative equilibria. The other equilibria involves the villagers not cooperating in any of their interactions. We will call this the non-cooperative equilibria. The cooperative equilibria Pareto dominates the non-cooperative equilibria. As such, we might expect it to be a more likely outcome of this static village economy game.[4] Note that, both of these equilibria are symmetric equilibria. Furthermore, regardless of which equilibrium is chosen in this two-person static village economy, the common-property resource will be over-exploited.

4.2.2 A three person example of the village economy game

The common property resource interaction in each period is represented by the stage game in Table 4.5, where $M > W > Y > N > Z > X > 0$ and $3W > M + 2N$:

Table 4.5: Three player common property resource interaction

	C_3		D_3	
	C_2	D_2	C_2	D_2
C_1	W, W, W	N, M, N	N, N, M	X, Y, Y
D_1	M, N, N	Y, Y, X	Y, X, Y	Z, Z, Z

[4] Pareto dominance is one of the two equilibrium selection criteria that were proposed by by Harsanyi and Selten ([47], pp. 355-357). They called this criterion payoff dominance. The other criterion that they proposed was risk dominance.

There are n other interactions between the villagers in each period. Each of these interactions can be represented by the stage game in Table 4.6, where $2V > C > V > 0$:

Table 4.6: Three player other interactions

	C_3		D_3	
	C_2	D_2	C_2	D_2
C_1	$V-\frac{C}{3}$ $V-\frac{C}{3}$ $V-\frac{C}{3}$	$V-\frac{C}{2}$ 0 $V-\frac{C}{2}$	$V-\frac{C}{2}$ $V-\frac{C}{2}$ 0	$V-C$ 0 0
D_1	0 $V-\frac{C}{2}$ $V-\frac{C}{2}$	0 0 $V-C$	0 $V-C$ 0	0 0 0

Just as in the two-person example, there are two pure strategy Nash equilibria in the three-person version of this village economy game. Once again, these are both symmetric equilibria, one of which involves cooperation among the villagers and one of which does not involve cooperation among the villagers. The cooperative equilibria involves the villagers cooperating in the provision of all of the public goods, but not cooperating in the exploitation of the common property resource. The non-cooperative equilibria involves the villagers not cooperating in any of their interactions. The cooperative equilibria Pareto dominates the non-cooperative equilibria, so that we might expect it to be a more likely outcome of this static village economy game. Nonetheless, regardless of which equilibrium is chosen in this three-person static village economy, the common-property resource will be over-exploited.

4.3 Outcomes in a village economy

Imagine a community that repeatedly plays the village economy game outlined in the previous section an infinite number of times. We will assume throughout that all villagers maximise the discounted presented value of their infinite sequence of per-period expected utilities. These per-period expected utilities are given by the payoffs from the village economy stage game described in the previous section of this paper. Furthermore, all villagers have a common one-period discount factor, $\delta \in (0, 1)$, that does not vary over time. Furthermore, we will assume that following the completion of the stage game in each period, every villager observes the outcomes of all of the activities in which they interact with the other villagers. We are interested in the impact that multiple interactions, both over time and across activities, have on the ability of a village to sustain cooperation in the use of a common property resource. This requires some benchmarks, against which we can compare the outcomes that occur when both types of multiple interaction are present. As such, we first consider what would happen in the absence of multiple interactions between villagers. This includes situations in which there are no multiple interactions in any dimension, as well as the situations in which the multiple interactions only take place either across activities or over time.

4.3.1 The outcome for a short-run association of hermits

Imagine a community whose members only interact through their use of the common property resource. We will call such a community an association of hermits, reflecting the solitary nature of its members. In this section, we consider a situation in which the common property interaction only takes place once. This short-run association of hermits is the most extreme of the benchmark scenarios.

Proposition 45 *A short-run association of hermits cannot avoid a tragedy of the commons without the use of private property or some other form of explicit regulation.*

Proof. A short-run association of hermits only plays the common property resource interaction component of the stage game. Furthermore, they only play it once. Note that $M > W$ and $Z > X$. As such, non-cooperation (D_i) is a strictly dominant strategy for each villager in the common property resource game. The unique Nash equilibrium in this game involves all villagers choosing not to cooperate. This means that the common property resource will be overused in the absence of some explicit regulatory regime that alters the payoffs in the common property resource game. ∎

4.3.2 Outcomes for a short-run village

Now imagine a community whose members interact in many activities. We will call such a community a village, reflecting the somewhat gregarious nature of its members. Suppose that each of these interactions only take place once and occur at a single point in time. This short-run village provides a benchmark in which the members of a community have multiple interactions across activities but not over time. This restriction means that villagers cannot link their behaviour in one activity to the outcome of another activity.

Proposition 46 *Short-run villages cannot avoid a tragedy of the commons without the use of private property or some other form of explicit regulation.*

Proof. A short-run village only plays the stage game once. Furthermore, the outcome of the common property interaction cannot be observed before the villagers choose their actions in the public goods interactions. As such, deviation in the common property resource interaction cannot be punished in these other interactions. This means that we need only consider the optimal behaviour of each villager in the common property interaction. In effect, we can treat the common property interaction as a separate game. We proved in Proposition 45 that non-cooperation (D_i) is a strictly dominant strategy for each villager in the common property resource game. Thus the unique Nash equilibrium in this game involves all villagers choosing not to cooperate. This means that the common property resource will be overused in the absence of some explicit regulatory regime that alters the payoffs in the common property resource game. ■

4.3.3 Outcomes for a long-run association of hermits

The final benchmark scenario involves a community of individuals who only interact through their use of a common property resource but do so repeatedly over time. Such a community will be called a long-run association of hermits. We will assume that there are an infinite number of repetitions of the common property game. The fact that the interactions take place repeatedly over time means that the hermits can link their behaviour in any interaction to the outcome of previous interactions.

Proposition 47 *A long-run association of hermits can avoid a tragedy of the commons if every hermit has a discount factor that is no smaller than* $\widehat{\delta}_0 = \frac{M-W}{M-Z} \in (0,1)$.

Proof. Hermits only interact through their use of the common property resource. As such, a long-run association of hermits involves infinite repetition of the common-property resource stage game. The efficient outcome in the stage game involves every hermit choosing to cooperate in the use of the common property resource. Suppose that the hermits attempt to sustain this outcome through the use of Nash reversion grim trigger strategies. The maximum net gain to any hermit who deviates from this strategy is:

$$Net\text{-}Gain_i(Deviation) = (M - W) - \sum_{t=1}^{\infty} \delta_i^t (W - Z).$$

If a tragedy of the commons is to be avoided, we need this net gain to be non-positive. This requires that:

$$\delta_i \geq \widehat{\delta}_0 = \frac{M - W}{M - Z} \text{ for all } i \in \{1, 2, \cdots, I\}.$$

Note that $M > W > Z > 0$, so that $(M - Z) > (M - W) > 0$. This ensures that $\widehat{\delta}_0 \in (0, 1)$. ■

4.3.4 Outcomes for a long-run village

A long-run village is a community whose members interact with each other both across activities and over time. We have already seen that interaction across activities alone does not allow a community to avoid a tragedy of the commons, but that interaction across time alone might do so. We now look at the impact that interaction across many activities has on the ability to avoid a tragedy of the commons for a community whose members also interact across time. We will assume that the short-run village stage game is repeated an infinite number of times.

Proposition 48 *A long-run village can avoid a tragedy of the commons if every villager has a discount factor that is no smaller than* $\widehat{\delta}_n = \left(\frac{M-W}{M-Z+n(V-E)}\right) \in (0,1)$*, where* n *is the number of interactions between the villagers that do not involve the common property resource.*

Proof. Suppose that every villager except for villager i employs the following Nash-reversion grim trigger strategy. They will cooperate in all interactions initially and will continue to do so as long as every other villager has also cooperated in all prior interactions. However, if any villager ever chooses not to cooperate, they will revert to playing the bad Nash equilibrium strategies in the short-run village stage game. This involves non-cooperation in all interactions. If villager i deviates from cooperation in any period, he will do so by overusing the common property resource. He gains nothing by deviating in the other interactions and still incurs the punishment from the rest of the village. As such, the gain to deviating for villager i is:

$$Benefit_i(deviate) = M - W.$$

If villager i deviates, then he will also incur a cost during the punishment phase of the super-game. This cost is given by:

$$Cost_i(deviate) = \delta_i \left(\sum_{t=0}^{\infty} \delta_i^t \left[(W - Z) + n \left\{ \left(V - \frac{C}{I} \right) - 0 \right\} \right] \right),$$

which can be simplified to yield:

$$Cost_i(deviate) = \frac{\delta_i \left\{ W - Z + n \left(V - \frac{C}{I} \right) \right\}}{(1 - \delta)}.$$

Clearly, villager i will deviate if and only if:

$$Benefit_i(deviate) - Cost_i(deviate) > 0.$$

This requires that:

$$\delta_i < \widehat{\delta}_n = \left(\frac{M - W}{M - Z + n \left(V - \frac{C}{I} \right)} \right).$$

In other words, the long-run village can deter deviation by a villager whenever he is sufficiently patient ($\delta_i \geq \widehat{\delta}_n$). Note that $M > W > Z > 0$, $2V > C > 0$ and $I \geq 2$. As such, we know that $(M - Z) > (M - W) > 0$ and $\left(V - \frac{C}{I}\right) > 0$. This ensures that $\widehat{\delta}_n \in (0, 1)$ for all $n \geq 0$. Furthermore, since this game is symmetric and the threatened punishment involves villagers choosing Nash-equilibrium strategies in the stage game, cooperation in all interactions can be sustained as a sub-game perfect equilibrium of the repeated game if all villagers are sufficiently patient. Specifically, this requires that $\delta_i \geq \widehat{\delta}_n$ for all $i \in \{1, 2, \cdots, I\}$. ■

We have established the circumstances under which a long-run village can avoid a tragedy of the commons. Recall that villagers interact with each other more times in each period than hermits. This difference results in villagers being able to avoid a tragedy of the commons in a larger set of circumstances than hermits.

Proposition 49 *A long run village can avoid a tragedy of the commons for a larger range of discount factors than a long-run association of hermits with the same number of members.*

Proof. We have

$$\widehat{\delta}_n - \widehat{\delta}_0 = \frac{-n\left(M - W\right)\left(V - \frac{C}{I}\right)}{\left(M - Z\right)^2 + n\left(M - Z\right)\left(V - \frac{C}{I}\right)}.$$

Note that $M > W > Z > 0$, $2V > C > 0$, $I \geq 2$ and $n > 0$. This ensures that $\min\left\{\left(M - W\right), \left(M - Z\right), \left(V - \frac{C}{I}\right), n\right\} > 0$. Hence we can conclude that $\widehat{\delta}_n - \widehat{\delta}_0 < 0$, so that $\widehat{\delta}_n < \widehat{\delta}_0$. ■

Finally, it is worth noting that the threshold discount factor for the successful management of the local common, $\widehat{\delta}_n$, falls as the number of contemporaneous interactions between villagers rises.

Proposition 50 *As the number of contemporaneous interactions between the* *fixed number of villagers rises, the threshold discount factor falls. Furthermore,* $\widehat{\delta}_n \to 0$ *as* $n \to \infty$.

Proof. If we differentiate $\widehat{\delta}_n$ with respect to n, we obtain

$$\frac{\partial \widehat{\delta}_n}{\partial n} = \frac{\partial \left\{ (M - W) \left[M - Z + n (V - C) \right]^{-1} \right\}}{\partial n} = \frac{-(M - W)\left(V - \frac{C}{I}\right)}{\left[M - Z + n\left(V - \frac{C}{I}\right)\right]^2}.$$

Note that $M > W > Z > 0$, $2V > C > 0$ and $I \geq 2$. As such, we know that $(M - Z) > 0$ and $\left(V - \frac{C}{I}\right) > 0$, so that $\left[M - Z + n(V - C)\right]^2 > 0$. Furthermore, it also ensures that $(M - W) > 0$, so that $\left\{-(M - W)\left(V - \frac{C}{I}\right)\right\} < 0$. Thus we can conclude that

$$\frac{\partial \widehat{\delta}_n}{\partial n} < 0.$$

Furthermore,

$$\lim_{n \to \infty} \widehat{\delta}_n = \lim_{n \to \infty} \left(\frac{M - W}{M - Z + n(V - E)} \right) = \left(\frac{M - W}{\infty} \right) = 0.$$

As such, we know that as n rises, $\widehat{\delta}_n$ must be falling towards zero. ∎

In other words, the probability that a long-run village with a fixed population can avoid a tragedy of the commons increases as the number of contemporaneous interactions between villagers increases. Note, however, that this result is contingent on the fact that the villagers are already interacting with each other repeatedly over time. The contemporaneous interactions enhance the effectiveness of the temporal interactions.

4.4 The impact of economic development

As economic development takes place, this village is likely to develop contacts with the outside world. The opportunity for trade with the outside world is likely to reduce the number of interactions between villagers. Goods and services can now be purchased from outside communities where their production previously required cooperation among the villagers. As such, the process of development can be incorporated in the simple model of a village economy through a reduction in the number of interactions among villagers that do not involve the common property resource. In order to draw direct inferences from the results we derived in the previous section, it is necessary to make some simplifying assumptions. First, we need to assume that the membership of the village does not change as the process of development takes place. Second, we need to assume that villagers hold static expectations about the number of contemporaneous interactions between them in future periods. These assumptions allow us to use the results from the previous section of this paper to draw some inferences about the impact of economic development on the ability of the village to successfully manage the local common.

Initially, when the village is a small and closed economy, the number of interactions between villagers is likely to be rather large. As such, it is highly probable that the village could have deterred individual villagers from over-using the common property resource. In the model, this takes the form of a relatively low threshold discount rate. However, as the village develops and increases its interaction with the outside world, the number of interactions among villagers falls. This leads to an increase in the threshold discount rate.

Proposition 51 *As the number of contemporaneous interactions between the fixed number of villagers rises, the threshold discount factor falls. Furthermore,* $\widehat{\delta}_n \to \widehat{\delta}_0$ *as* $n \to 0$*.*

Proof. In Proposition 49, we established that $\widehat{\delta}_n < \widehat{\delta}_0$. Furthermore, in the process of proving Proposition 50, we established that $\widehat{\delta}_n$ is a strictly decreasing function of n. Finally, note that

$$\lim_{n\to 0}\widehat{\delta}_n = \lim_{n\to 0}\left(\frac{M-W}{M-Z+n\left(V-E\right)}\right) = \left(\frac{M-W}{M-Z}\right) = \widehat{\delta}_0.$$

Since $\widehat{\delta}_n$ is a strictly decreasing function of n that approaches $\widehat{\delta}_0$ as $n \to 0$, we can conclude that, as n falls, $\widehat{\delta}_n$ must be rising towards $\widehat{\delta}_0$. ■

Eventually, there may come a point where the village can no longer deter individual villagers from over-using the common property resource. At that point, some other solution will be required if a tragedy of the commons is to be avoided. This solution might involve some form of explicit regulation or it might involve the allocation of private property rights.

Proposition 52 *If a long-run association of hermits cannot avoid a tragedy of the commons, then there exists some minimum number of contemporaneous interactions between villagers, $\widehat{n}(\delta) > 0$, that must occur if a long-run village whose members have the same common discount factor as the hermits is to avoid a tragedy of the commons.*

Proof. If a long-run association of hermits cannot avoid a tragedy of the commons, then the hermits must be too impatient. This means that

$$\delta < \frac{M-W}{M-Z},$$

so that

$$(M-W) - \delta(M-Z) > 0.$$

Suppose that a long-run village can avoid a tragedy of the commons. Then it must be the case that $\delta \in \left[\widehat{\delta}_n, \widehat{\delta}_0\right)$. This requires that $\delta \geq \widehat{\delta}_n$, so that

$$\delta \geq \left(\frac{M-W}{M-Z+n\left(V-\frac{C}{I}\right)}\right).$$

This inequality can be rearranged to obtain

$$n \geq \widehat{n}(\delta) = \left(\frac{(M - W) - \delta(M - Z)}{\delta \left(V - \frac{C}{I}\right)} \right).$$

Since $(M - W) - \delta(M - Z) > 0$, $\delta \in (0, 1)$, $2V > C > 0$ and $I \geq 2$, we know that $\widehat{n}(\delta) > 0$. ■

Note that in this simple model of a village economy, rather than private property rights being a precondition for development to take place, it is the process of development that leads to the need for private property rights or some other mechanism to manage local commons. The cost of implementing a more explicit regime for managing the local common is a hidden cost of development. This does not necessarily mean, however, that economic development reduces social welfare. The benefits from economic development may well exceed the costs of developing and implementing an explicit regime for managing common property resources.

4.5 Conclusion

In this chapter, we have provided a simple explanation for the observation that small, isolated communities can sometimes manage common property resources effectively without recourse to private property rights or other forms of explicit regulation. The explanation requires the villagers to interact with each other in a number of activities. It also requires these interactions to be repeated over time. This non-anonymous multi-market contact enhances the ability of the village to deter individual villagers from over-using the common property resource. This insight is similar to the idea that multi-market contact might enhance the ability of firms to sustain collusion.[5] However, the process of economic development is

[5] See, for example, Bernheim and Whinston ([13]).

likely to reduce the need for villagers to directly interact with each other. As the number of activities in which villagers interact declines, it becomes harder for the village to deter over-use of the common property resource. Eventually, an alternative regime for managing the common property resource will be needed. This might involve the allocation of explicit private property rights or other legal restrictions on the use of the common property resource. The cost of implementing such a system is, in a sense, a hidden cost of development.

There are a number of potential extensions to the model developed in this paper. First, the static nature of the villagers' beliefs about the number of interactions could be relaxed. A first step in this direction would involve the specification of an exogenous stochastic process that governs the evolution of the number of such interactions. If we also relaxed the assumption of non-durability of the stock of each public good, the evolution of the number of interactions could be endogenised. This would be accomplished by making the probability of a reduction in the number of interactions a function of the existing stock of one or more of the local public goods. Another potential extension to the model would involve replacing the static version of the tragedy of the commons that is employed in this paper with a dynamic version of the tragedy of the commons. Finally, it would be worth exploring the relationship between village size and the number of non-anonymous transactions between the various villagers. In general, the larger a community becomes, the less likely it will be that each villager knows and directly interacts with every other villager. As such, the relationship between community size and the number of non-anonymous interactions is likely to be non-linear. Initially, an increase in the village population might increase the number of non-anonymous transactions. However, eventually a point might be reached at which further increases in village population result in a decrease in the number of non-anonymous transactions among villagers.

4.6 Appendix: A formal model of a village economy

Consider a village in which there are I individuals, indexed by $i \in \{1, 2, \cdots, I\}$. Initially, the village has no interaction with the outside world. As such, it constitutes a closed economy. This assumption will be relaxed later. Each of the villagers uses his own labour (N_i) and some amount of a common property resource (K_i) to produce a private consumption good for his own consumption. Villager i's production of this private consumption good is denoted by c_i. In addition to this, the villagers pool their labour to produce some commodities that take the form of excludible public goods. There are J of these public commodities. They are indexed by $j \in \{1, 2, \cdots, J\}$. Both the private and public goods are perishable and cannot be stored over time. We will assume that all villagers have identical preferences that can be represented by a constant discount rate, δ, and a per-period Bernoulli utility function, $U : \mathbb{R}_+ \times \{0, 1\}^J \rightarrow \mathbb{R}$. This utility function is defined over a villager's consumption of leisure, the private good and each of the J excludible public goods, $U(l_i, c_i, q_1, q_2, \cdots, q_J)$. For simplicity, assume that each villager's per-period Bernoulli utility function is additively separable. As such, we can write $U(l_i, c_i, q_1, q_2, \cdots, q_J) = u_l(l_i) + u_c(c_i) + \sum_{j=1}^{J} u_j(q_j)$.

Each of the private goods is produced using a production technology, $F_i : \mathbb{R}_+^{I+1} \rightarrow \mathbb{R}_+$, that can be represented by a production function of the form $y_i = F(N_i; N_{-i}, \overline{K})$, where $\overline{K}$ is the total stock of the common property resource in every period. To keep things simple, we will assume that this production technology is the same for every villager, $i \in \{1, 2, \cdots, I\}$. The amount of the consumption good that is produced by a particular villager in any given period depends in part on the total usage of the resource by the village in that period. Specifically, $c_n = F(N_i; N_{-i}, \overline{K}) = \frac{N_{n,J+1}}{N_{J+1}} \widehat{F}(N_{j+1}; \overline{K}) = f(N_{j+1})$, where $N_{J+1} = \sum_{i=1}^{I} N_{i,J+1}$, $f'(N_{J+1}) > 0$ and $f''(N_{J+1}) < 0$.

Each of the public goods is produced using a production technology, $g_j : \mathbb{R}_+ \rightarrow \{0, 1\}$, which can be represented by a production function of the form $q_j = g_j(N_j)$. Here $q_j \in \{0, 1\}$, where $q_j = 1$ means that the public good is produced and $q_j = 0$ means that it is not produced. Note that the public goods are all or nothing commodities. They are either produced or they are not produced. The labour input to the production of each public good is the aggregate amount of labour devoted to its production by the entire village, so that $N_j = \sum_{i=1}^{I} N_{i,j}$. There is some threshold level of labour that is required to produce a particular public good. This threshold level of labour is denoted by $\widehat{N}_j$.

Labour is in scarce supply. Thus the village economy faces a number of feasibility constraints in each period. Each villager has $\overline{N}$ units of labour to allocate between various activities in each period. As such, the labour feasibility constraints take the form $\sum_{j=0}^{J+1} N_{i,j,t} \leq \overline{N}$ for all $i \in \{1, 2, \cdots, I\}$. Here, $j = 0$ denotes the amount of time that a villager devotes to leisure, while $j = J + 1$ denotes the amount of time a villager devotes to producing the private consumption good.

Bibliography

[1] Abreu, D, D Pearce and E Stacchetti (1987), "Optimal cartel equilibria with imperfect monitoring", *Journal of Economic Theory 39(1)*, pp. 251-269.

[2] Abreu, D, D Pearce and E Stacchetti (1990), "Toward a theory of discounted repeated games with imperfect monitoring", *Econometrica 58(5)*, pp. 1041-1063.

[3] Akerlof, GA (1970), "The market for "lemons": Quality uncertainty and the market mechanism", *Quarterly Journal of Economics 84(3)*, August, pp. 488-500.

[4] Alchian, AA and WR Allen (1983), *Exchange and production: Competition, coordination and control (third edition)*, Wadsworth Publishing Company, USA.

[5] Arrow, KJ (1951), "An extension of the basic theorems of classical welfare economics", pp. 507-532 in Neyman, J (Editor) (1951), *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Held at the Statistical Laboratory of the Department of Mathematics at the University of California between 31 July 1950 and 12 August 1950, University of California Press, Berkeley and Los Angeles, California, USA.

Reprinted as Chapter 2 (pp. 13-45) in Arrow, KJ (1983), *Collected papers of Kenneth J Arrow — Volume 2: General equilibrium*, Basil Blackwell, USA.

[6] Arrow, KJ (1963), "Uncertainty and the welfare economics of medical care", *American Economic Review 53(5)*, pp. 941-973.

[7] Arrow, KJ (1965), "Uncertainty and the welfare economics of medical care: Reply (The implications of transaction costs and adjustment lags)", *American Economic Review 55(1 and 2)*, March-May, pp. 154-158.

[8] Arrow, KJ (1968), "The economics of moral hazard: Further comment", *American Economic Review 58(3)(Part 1)*, June, pp. 537-539.

[9] Arrow, KJ and G Debreu (1954), "Existence of an equilibrium for a competitive economy", *Econometrica 22(3)*, July, pp. 265-290. Reprinted as Chapter 4 (pp. 58-91) in Arrow, KJ (1983), *Collected papers of Kenneth J Arrow — Volume 2: General equilibrium*, Basil Blackwell, USA. Also reprinted as Chapter 4 (pp. 68-97) in Debreu, G (1983), *Mathematical economics: Twenty papers of Gerard Debreu*, With an introduction by Werner Hildenbrand, Econometric Society Monographs in Pure Theory, Econometric Society Publications Number 4, Cambridge University Press, USA.

[10] Atkeson, A and R Lucas (1992), "On efficient distribution with private information", *Review of Economic Studies 59(200)*, pp. 427-453.

[11] Banks, JS and J Sobel (1987), "Equilibrium selection in signaling games", *Econometrica 55(3)*, May, pp. 647-661.

[12] Bardhan, P (1993), "Symposium on management of local commons", *Journal of Economic Perspectives 74(4)*, Fall, pp. 87-92.

[13] Bernheim, BD and MD Whinston (1990), "Multimarket contact and collusive behavior", *Rand Journal of Economics 21(1)*, Spring, pp. 1-26.

[14] Biglaiser, G (1993), "Middlemen as experts", *Rand Journal of Economics 24(2)*, pp. 212-223.

[15] Biglaiser, G and JW Friedman (1994), "Middlemen as guarantors of quality", *International Journal of Industrial Organization 12(4)*, pp. 509-531.

[16] Billingsley, P (1995), *Probability and measure (third edition)*, John Wiley and Sons, USA.

[17] Bolton, P and M Dewatripont (2005), *Contract theory*, Massachusetts Institute of Technology Press, USA.

[18] Border, KC (2000), "Brief notes on the Arrow-Debreu-McKenzie model of an economy", mimeo California Institute of Technology, USA, January.

[19] British Medical Association (2007), "NHS hospital treatment and information about specialists (updated June 2007)", Information available online at (http://www.bma.org.uk/ap.nsf/Content/infospecialists). Downloaded on 6 November 2007.

[20] Cho, IK and DM Kreps (1987), "Signaling games and stable equilibria", *Quarterly Journal of Economics 102(2)*, May, pp. 179-222.

[21] Coase, RH (1960), "The problem of social cost", *Journal of Law and Economics 3*, October, pp. 1-44.

[22] Commonwealth of Australia (2005), *General practice in Australia: 2004*, GP Communications and Business Improvement Unit, Budget and Performance Branch, Primary Care Division, Department of Health and Ageing, Canberra, May (ISBN: 0-642-82-676-5).

[23] Commonwealth of Australia (Undated), "Medicare welcome kit (English version)", available online as a portable document format file at (http://www.medicareaustralia.gov.au/resources/welcome_kits/english/ma_welcome_kit_english_medicare.pdf). Downloaded on 6 November 2007.

[24] Cooper, R and TW Ross (1984), "Prices, product qualities and asymmetric information: The competitive case", *Review of Economic Studies 51(2)*, pp. 197-207.

[25] Cornes, R and T Sandler (1996), *The theory of externalities, public goods and club goods (second edition)*, Cambridge University Press, USA.

[26] Damania, R and D Round (2000), "The economics of consumer protection: Introduction", *Australian Economic Papers 39(4)*, pp. 403-407.

[27] Debreu, G (1951), "The coefficient of resource utilzation", *Econometrica 19(3)*, July, pp. 273-292. Reprinted as Chapter 1 (pp. 30-49) in Debreu, G (1983), *Mathematical economics: Twenty papers of Gerard Debreu*, With an introduction by Werner Hildenbrand, Econometric Society Monographs in Pure Theory, Econometric Society Publications Number 4, Cambridge University Press, USA.

[28] Debreu, G (1954), "Valuation equilibrium and Pareto optimum", *Proceedings of the National Academy of Sciences 40(7)*, July, pp. 588-592. Reprinted as Chapter 5 (pp. 98-104) in Debreu, G (1983), *Mathematical economics: Twenty papers of Gerard Debreu*, With an introduction by Werner Hildenbrand, Econometric Society Monographs in Pure Theory, Econometric Society Publications Number 4, Cambridge University Press, USA.

[29] Debreu, G (1959), *Theory of value: An axiomatic analysis of economic equilibrium*, Cowles Foundation Monograph Number 17, Cowles Foundation for Research in Economics at Yale University, Yale University Press, USA.

[30] De Long, JB and A Shleifer (1993), "Princes and merchants: European city growth before the industrial revolution", *Journal of Law and Economics 36(2)*, October, pp. 671-702.

[31] Diamond, DW (1989), "Reputation acquisition in debt markets", *Journal of Political Economy 97(4)*, pp. 828-862.

[32] Dietz, T, N Dolsak, E Ostrom and PC Stern (2002), "The drama of the commons", Chapter 1 (pp. 3-35) in Ostrom, E, T Dietz, N Dolsak, PC Stern, S Stonich and EU Weber (Editors) (2002), *The drama of the commons*, Committee on the Human Dimensions of Global Change, Division of Behavioral and Social Sciences and Education, National Research Council, National Academy Press, USA.

[33] Dietz, T, E Ostrom and PC Stern (2003), "The struggle to govern the commons", *Science 302*, 12 December, pp. 1907-1912.

[34] Dixit, A (1987), "Trade and insurance with moral hazard", *Journal of International Economics 23(3 and 4)*, November, pp. 201-220.

[35] Dixit, A (1989), "Trade and insurance with adverse selection", *Review of Economic Studies 56(2)*, April, pp. 235-247.

[36] Ehrlich, I and GS Becker (1972), "Market insurance, self-insurance and self-protection", *Journal of Political Economy 80(4)*, July-August, pp. 623-648.

[37] Ely, JC, J Horner and W Olszewski (2005), "Belief-free equilibria in repeated games", *Econometrica 73(2)*, pp. 377-415.

[38] Folland, S, AC Goodman and M Stano (1993), *The economics of health and health care (second edition)*, Prentice-Hall, USA.

[39] Fudenberg, D, D Levine and E Maskin (1994), "The folk theorem with imperfect public information", *Econometrica 62(5)*, pp. 997-1039.

[40] Fudenberg, D and J Tirole (1995), *Game theory*, Massachusetts Institute of Technology Press, USA.

[41] Gordon, HS (1954), "The economic theory of a common property resource: The fishery", *Journal of Political Economy 62(2)*, April, pp. 124-142.

[42] Grafton, RQ, D Squires and KJ Fox (2000), "Private property and economic efficiency: A study of a common-pool resource", *Journal of Law and Economics 43(2)*, October, pp. 679-713.

[43] Green, EJ and RH Porter (1984), "Noncooperative collusion under imperfect price information", *Econometrica 52(1)*, pp. 87-100.

[44] Grossman, S and O Hart (1983), "An analysis of the principal-agent problem", *Econometrica 51(1)*, pp. 7-45.

[45] Hadfield, GK, R Howse and MJ Trebilcock (1998), "Information based principles for rethinking consumer protection policy", *Journal of Consumer Policy 21(2)*, pp. 131-169.

[46] Hardin, G (1968), "The tragedy of the commons", *Science (New Series) 62(3859)*, 13 December, pp. 1243-1248.

[47] Harsanyi, JC and R Selten (1988), *A general theory of equilibrium selection in games*, Massachusetts Institute of Technology Press, USA.

[48] Hermalin, BE and ML Katz (1991), "Moral hazard and verifiability: The effects of renegotiating in agency", *Econometrica 59(6)*, pp. 1735-1754.

[49] Hirshleifer, J and J Riley (1992), *The analytics of uncertainty and information*, Cambridge Surveys of Economic Literature, Cambridge University Press, Great Britain.

[50] Hogan, C (2001), "Enforcement of implicit employment contracts through unionization", *Journal of Labor Economics 19(1)*, pp. pp. 171-195.

[51] Hogg, RV and AT Craig (1978), *Introduction to mathematical statistics (fourth edition)*, Macmillan, USA.

[52] Holmstrom, B (1979), "Moral hazard and observability", *Bell Journal of Economics 10(1)*, pp. pp. 74-91.

[53] Holmstrom, B (1999), "Managerial incentive problems: A dynamic perspective", *Review of Economic Studies 66(1)*, pp. 169-182.

[54] Horner, J (2002), "Reputation and competition", *American Economic Review 92(3)*, pp. 644-663.

[55] Jewitt, I (1988), "Justifying the first-order approach to principal-agent problems", *Econometrica 56(5)*, pp. 1177-1190.

[56] Johnson, S, J McMillan and C Woodruff (2002), "Property rights and finance", *American Economic Review 92(5)*, December, pp. 1335-1356.

[57] Judd, KL (1985), "The law of large numbers with a continuum of iid random variables", *Journal of Economic Theory 35(1)*, February, pp. 19-25.

[58] Klein, B and KB Leffler (1981), "The role of market forces in assuring contractual performance", *Journal of Political Economy 89(4)*, August, pp. 615-641.

[59] Kocherlakota, NR (1996), "Implications of risk sharing without commitment", *Review of Economic Studies 63(4)*, pp. 595-609.

[60] Kreps, DM (1990a), *A course on microeconomic theory*, Harvester Wheatsheaf, Great Britain.

[61] Kreps, DM (1990b), "Corporate culture and economic theory", Chapter 4 (pp. 90-143) in Alt, JE and KA Shepsle (Editors) (1990), *Perspectives on positive political economy*, Cambridge University Press, USA.

[62] Laffont, JJ and D Martimort (2002), *The theory of incentives: The principal-agent model*, Princeton University Press, USA.

[63] Lees, DS and RG Rice (1965), "Uncertainty and the welfare economics of medical care: Comment", *American Economic Review 55(1 and 2)*, March-May, pp. 140-154.

[64] Macho-Stadler, I and JD Perez-Castrillo (2001), *An introductrion to the economics of information: Incentives and contracts (second edition)*, Oxford University Press, Great Britain.

[65] MacLeod, B and JM Malcolmson (1989), "Implicit contracts, incentive compatibility and involuntary unemployment", *Econometrica 57(2)*, pp. 447-480.

[66] Mailath, GJ and L Samuelson (2006), *Repeated games and reputations: Long-run relationships*, Oxford University Press, USA.

[67] Malcolmson, JM (1983), "Trade unions and economic efficiency", *Economic Journal 93 (Supplement: Conference Papers)*, pp. 51-65.

[68] Mas-Colell, A, MD Whinston and JR Green (1995), *Microeconomic theory*, Oxford University Press, USA.

[69] McKenzie, LW (1954), "On equilibrium in Graham's model of world trade and other competitive systems", *Econometrica 22(2)*, April, pp. 147-161.

[70] McKenzie, LW (1959), "On the existence of general equilibrium for a competitive market", *Econometrica 27(1)*, January, pp. 54-71.

[71] McKenzie, LW (1961), "On the existence of general equilibrium: Some corrections", *Econometrica 29(2)*, April, pp. 247-248.

[72] Mirlees, JA (1999), "The theory of moral hazard and unobservable behaviour: Part 1", *Review of Economic Studies 66(1)*, pp. 3-21.

[73] Mooney, G and M Ryan (1993), "Agency in health care: Getting beyond first principles", *Journal of Health Economics 12(2)*, pp. 125-135.

[74] Murphy, KM, A Shleifer and RW Vishny (1993), "Why is rent-seeking so costly to growth?", *American Economic Review 83(2)*, Papers and Proceedings of the One Hundred and Fifth Annual Meeting of the American Economic Association, May, pp. 409-414.

[75] Myerson, R (1979), "Incentive compatibility and the bargaining problem", *Econometrica 47(1)*, pp. 61-73.

[76] Negishi, T (1960), "Welfare economics and the existence of equilibrium for a competitive economy", *Metroeconomica 12*, pp. 92-97.

[77] Nikaido, H (1956), "On the classical multilateral exchange problem", *Metroeconomica 8*, pp. 135-145.

[78] O'Rourke, PJ (2007), *On the wealth of nations*, Allen and Unwin, Australia.

[79] Ostrom, E, J Burger, CB Field, RB Norgaard and D Policansky (1999), "Revisiting the commons: Local lessons, global challenges", *Science (New Series) 284(5412)*, 9 April, pp. 278-282.

[80] Ostrom, E and R Gardner (1993), "Coping with asymmetries in the commons: Self-governing irrigation systems can work", *Journal of Economic Perspectives 74(4)*, Fall, pp. 93-112.

[81] Pauly, MV (1968), "The economics of moral hazard: Comment", *American Economic Review 58(3)(Part 1)*, June, pp. 531-537.

[82] Pauly, MV (1974), "Overinsurance and public provision of insurance: The roles of moral hazard and adverse selection", *Quarterly Journal of Economics 88(1)*, February, pp. 44-62.

[83] Powell-Davies, G and D Fry (2005), "General practice in the health system", Chapter 10 (pp. 420-464) of Commonwealth of Australia (2005), *General practice in Australia: 2004*, GP Communications and Business Improvement Unit, Budget and Performance Branch, Primary Care Division, Department of Health and Ageing, Canberra, May (ISBN: 0-642-82-676-5).

[84] Prescott, EC and RM Townsend (1984), "Pareto optima and competitive equilibria with adverse selection and moral hazard", *Econometrica 52(1)*, January, pp. 21-46.

[85] Prescott, ES and RM Townsend (2002), "Collective organizations versus relative performance contracts: Inequality, risk sharing and moral hazard", *Journal of Economic Theory 103(2)*, pp. 282-310.

[86] Pretty, J (2003), "Social capital and the collective management of resources", *Science 302*, 12 December, pp. 1912-1914.

[87] Radner, R (1982), "Monitoring cooperative agreements in a repeated principal-agent relationship", *Econometrica 49(5)*, pp. 1127-1148.

[88] Radner, R (1985), "Repeated principal-agent games with discounting", *Econometrica 53(5)*, pp. 1173-1198.

[89] Radner, R, R Myerson and E Maskin (1986), "An example of a partnership game with discounting and uniformly inefficient equilibria", *Review of Economic Studies 53(1)*, pp. 59-69.

[90] Riley, JG (1979), "Informational equilibrium", *Econometrica 47(2)*, March, pp. 331-359.

[91] Riley, JG (1985), "Competition with hidden knowledge", *Journal of Political Economy 93(5)*, October, pp. 958-976.

[92] Riley, JG (2001), "Silver signals: Twenty-five years of screening and signaling", *Journal of Economic Literature 39(2)*, pp. 432-478.

[93] Rogerson, W (1985a), "Repeated moral hazard", *Econometrica 53(1)*, pp. 69-76.

[94] Rogerson, W (1985b), "The first-order approach to principal-agent problems", *Econometrica 53(6)*, pp. 1357-1368.

[95] Ross, SA (1973), "The economic theory of agency: The principal's problem", *American Economic Review: Papers and Proceedings 63(2)*, pp. 134-139.

[96] Rothschild, M and J Stiglitz (1976), "Equilibrium in competitive insurance markets: An essay on the economics of imperfect information", *Quarterly Journal of Economics 90(4)*, November, pp. 629-649.

[97] Royal New Zealand College of General Practitioners (2005), *Your guide to New Zealand general practice*, Version 3, Royal New Zealand College of General Practitioners, New Zealand, July (ISBN: 0-9582429-1-7).

[98] Rubinstein, A and ME Yaari (1983), "Repeated insurance contracts and moral hazard", *Journal of Economic Theory 30(1)*, pp. 74-97.

[99] Samuelson, PA (1954), "The pure theory of public expenditure", *Review of Economics and Statistics 36(4)*, November, pp. 387-389.

[100] Samuelson, PA (1955), "Diagrammatic exposition of a theory of public expenditure", *Review of Economics and Statistics 37(4)*, November, pp. 387-389.

[101] Scully, GW (1988), "The institutional framework and economic development", *Journal of Political Economy 96(3)*, pp. 652-662.

[102] Seabright, P (1993), "Managing local commons: Theoretical issues in incentive design", *Journal of Economic Perspectives 7(4)*, Fall, pp. 113-134.

[103] Selten, R (1975), "Reexamination of the perfectness concept for equilibrium points in extensive games", *International Journal of Game Theory 4(1)*, March, pp. 25-55.

[104] Shapiro, C (1983), "Premiums for high quality products as returns to reputations", *Quarterly Journal of Economics 98(4)*, pp. 659-679.

[105] Shapiro, C and JE Stiglitz (1984), "Equilibrium unemployment as a worker discipline device", *American Economic Review 74(3)*, pp. 433-444.

[106] Shavell, S (1979a), "Risk sharing and incentives in the principal and agent relationship", *Bell Journal of Economics 10(1)*, pp. 55-73.

[107] Shavell, S (1979b), "On moral hazard and insurance", *Quarterly Journal of Economics 93(4)*, pp. 541-592.

[108] Simon, LK and MB Stinchcombe (1995), "Equilibrium refinements for infinite normal form games", *Econometrica 63(6)*, pp. 1421-1443.

[109] Smith, A (1759), *The theory of moral sentiments*, 1976 Clarendon Press Edition, Edited by DD Raphael and AL Macfie, Oxford University Press, Great Britain.

[110] Smith, A (1776a), *The wealth of nations: Books I-III*, 1999 Penguin Classics Edition, Edited with an introduction and notes by Andrew Skinner, Penguin Books, England.

[111] Smith, A (1776b), *The wealth of nations: Books IV-V*, 1999 Penguin Classics Edition, Edited with an introduction and notes by Andrew Skinner, Penguin Books, England.

[112] Smith, RL (2000), "When competition is not enough: Consumer protection", *Australian Economic Papers 39(4)*, pp. 408-425.

[113] Spear, SE and S Srivastava (1987), "On repeated moral hazard with discounting", *Review of Economic Studies 54(4)*, pp. 599-617.

[114] Spence, M (1973), "Job market signaling", *Quarterly Journal of Economics 87(3)*, August, pp. 355-374.

[115] Starr, RM (1997), *General equilibrium theory: An introduction*, Cambridge University Press USA.

[116] Stigler, GJ (1971), "Can regulatory agencies protect the consumer?", Chapter 11 (pp. 178-188) in Stigler, GJ (1975), *The citizen and the state: Essays on regulation*, University of Chicago Press, USA.

[117] Stiglitz, JE (1977), "Monopoly, non-linear pricing and imperfect information: The insurance market", *Review of Economic Studies 44(3)*, October, pp. 407-430.

[118] Stiglitz, JE (1979), "Equilibrium in product markets with imperfect information", *American Economic Review 69(2)*, Papers and Proceedings of the Ninety-First Annual Meeting of the American Economic Association, May, pp. 339-345.

[119] Stiglitz, JE and AE Weiss (1981), "Credit rationing in markets with imperfect information", *American Economic Review 7193)*, June, pp. 393-410.

[120] Stiglitz, J and A Weiss (1990), *Sorting out the differences between signaling and screening models*, NBER Technical Working Paper Number 93, National Bureau of Economic Research, USA.

[121] Tadelis, S (1999), "What's in a name? Reputation as a tradable asset", *American Economic Review 89(3)*, pp. 548-563.

[122] Takayama, A (1985), *Mathematical economics (second edition)*, Cambridge University Press, USA.

[123] Thomas, J and T Worrall (1990), "Income fluctuation and asymmetric information: An example of a repeated principal-agent problem", *Journal of Economic Theory 51(2)*, pp. 367-390.

[124] Tirole, J (1988), *The theory of industrial organization*, Massachusetts Institute of Technology Press, USA.

[125] Tirole, J (1996), "A theory of collective reputations (with applications to the persistence of corruption and to firm quality)", *Review of Economic Studies 63(1)*, pp. 1-22.

[126] Townsend, R (1982), "Optimal multiperiod contracts and the gain from enduring relationships under private information", *Journal of Political Economy 90(6)*, pp. 1166-1186.

[127] Wilson, C (1977), "A model of insurance markets with incomplete information", *Journal of Economic Theory 16(2)*, December, pp. 167-207.

Vita

Damien Sean Eldridge was born early in the morning on Friday 13 November 1970 at Manly Hospital in Sydney, Australia. He is the eldest son of Robyn Valerie Eldridge (nee Franklin) and William John Eldridge. Damien's preschool education was completed at North Curtin Preschool in Canberra. His primary school education was completed at Holy Trinity Primary School in Curtin (Kindergarten to Year Three) and Marist College Canberra (Year Four to Year Six). His secondary education (Year Seven to Year Twelve) was also completed at Marist College Canberra, from which he graduated in 1988. He began his tertiary education at the Australian National University in 1989 and went on to complete a Bachelor of Economics degree, a Bachelor of Science degree, a Graduate Diploma in Economics and a Master of Economics degree. Damien began graduate studies in economics at the University of Texas at Austin in August 1999. He completed a Master of Science degree at Texas in May 2001 and is in the process of completing a Doctor of Philosophy degree. In addition to his academic training, Damien has about six years of practical experience as an applied microeconomist. This experience was gained while working for a variety of Australian Public Service agencies and an economic consulting firm. He is currently an economics lecturer in the Department of Economics and Finance at La Trobe University. Damien's research interests include microeconomic theory and applied microeconomics.

Permanent address: PO Box 130, La Trobe University Post Office, Bundoora, Victoria, 3083, Australia.

This dissertation was typed by the author using the Scientific Workplace package developed by MacKichan Software. The figures were produced by the author using the Microsoft Word package developed by Microsoft.